

A Work Project, presented as part of the requirements for the Award of a Master's degree in
Business Analytics from the Nova School of Business and Economics.

SENTIMENT ANALYSIS-BASED PREDICTION OF “MEME STOCK” PRICE
MOVEMENTS

BENEDIKT NIKLAS BUSCH

Work project carried out under the supervision of:

Francisco Queiró

14/12/2022

Abstract

This paper investigates the relation of daily twitter sentiment to the stock prices of two portfolios for a three-year time-period, one consisting of popular meme stocks, the other of NASDAQ stocks. The objective is to detect differences in the behavior of both portfolios in this relation. In its sentiment classification, the paper makes use of Textblob, customized VADER, roBERTa as well as a manually trained GRU neural network while relying on classic binary classifiers to predict price movements. The results indicate a stronger predictability for the NASDAQ portfolio, most likely due to issues of information quality.

Keywords: sentiment analysis, meme stocks, twitter

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Table of Contents

1. Introduction	3
2. Literature Review	4
2.1. On the Emergence of “Meme Stocks”	4
2.2. Sentiment Analysis and Stock Correlation	6
3. Methodology	8
3.1. Data Mining and Extraction	8
3.2. Data Cleaning and Preparation	10
3.3. Model Selection and Preparation.....	11
3.3.1. Textblob	11
3.3.2. VADER.....	12
3.3.3. Gated Recurrent Unit	13
3.4. Merging Classifiers and Preparing Correlational Analysis	16
4. Evaluation.....	17
4.1. Time-lagged Cross-correlation	17
4.2. Binary Classification	20
4.3. Hypothetical Portfolio Strategy and Back Testing	21
4.4. Discussion.....	22
5. Limitations	23
6. Future Research.....	24
7. Conclusion.....	25
8. References	26
9. List of Figures and Tables	29
10. Appendices	30
Appendix A: Lagged Correlation Plots	30
Appendix B: Binary Classification Plots.....	33
Appendix C: Hypothetical Portfolio Back Testing.....	34

1. Introduction

The proliferation of social media has caused a plethora of unprecedented changes across the world. In particular, it has changed the way humanity consumes and interacts with the news. Over are the days when headlines were relayed daily via newspapers and other print media. Today, people consume news in a continuous, rapid, and on-demand fashion. Headlines are conveyed to smartphones instantly, and everyone is always connected, with Reuters research suggesting that 73% of people consume news using their smartphones (Newman et al. 2021). In 2020, the world was struck by a major pandemic, forcing everyone into quarantine, which further accelerated humanity's shift to digital. It is therefore no surprise that the financial markets and related communications have also moved to the internet. Hundreds of thousands of texts relating to financial securities and equities markets are sent and published by individuals and institutions daily, a significant portion of these on the popular social media platform Twitter as well as the forum Reddit. This spreading of securities-related information on platforms with huge user bases has made a massive and measurable impact on the way institutional and individual investors engage with the market. In January of 2021, the interrelation between financial markets and social media reached its zenith with the short squeeze of Gamestop Inc. While this was not the first ever short squeeze – Volkswagen / Porsche (Godfrey 2016) serve as a poster child of this mechanism – the story of Gamestop is unprecedented in that it involved the organization of mainly retail investors on social media platforms (Allen et al. 2021). This anomaly in stock market behavior caused a spillover effect to other highly shorted stocks with the hope of retail investors being to achieve similar short squeezes. Individuals attempted to generate hype on e.g., Twitter to drive stock prices up. In contrast to the assumptions of the efficient market hypothesis (EMH) and its flagship capital asset pricing model (CAPM), it seems probable that public sentiment or mood regarding said securities hereafter referred to as “meme stocks”, should show considerable levels of correlation with their price. Additionally,

given said relation, it would be interesting to investigate these securities' co-movement with sentiment in comparison with a selection of blue-chip companies. This thesis will thus attempt to answer the following research question:

Is the movement of meme stock prices predictable using social media sentiment and how does the predictability differ from traditional blue-chip stocks?

In answering this research question, first, this paper will conduct a critical review of existing literature from both machine learning and finance domains. Next, the paper will attempt to create a mechanism to measure social media interaction and public sentiment regarding chosen securities mined through Twitter by using pre-trained natural language processing (NLP) algorithms as well as by constructing a manually trained deep recurrent neural network (dRNN). Finally, this thesis will measure and compare the predictability of price movements in meme stocks versus traditional blue-chip stocks using sentiment.

2. Literature Review

2.1. On the Emergence of “Meme Stocks”

Central to this thesis is the discussion of the remarkable dynamics regarding the so-called meme stocks and how they differ from typical blue-chip stocks. It is therefore critical to define meme stocks and determine a way to differentiate these from traditional stocks. Especially in the scope of sentiment analysis based on social media, a chicken/egg problem can occur where it is unclear whether or not increased volatility creates social media hype which then christens the stock meme-worthy, or if the opposite is the case.

While meme stocks have long existed in the microcosm of the microblogging platform Reddit's

“wallstreetbets” subreddit, the Gamestop short squeeze of January 2021 catapulted the term into mainstream popularity. Similarly, short squeezes have also existed historically, but as posited by Allen et al. (2021), the Gamestop episode was unprecedented due to the close coordination among involved traders. According to the authors, this level of coordination was mainly made possible by the popularization of social media in recent years. In response, citing reasons of maintaining market efficiency, major trading platforms began levying trade restrictions against the stocks in question (Fitzgerald 2021). Allen et al. use this set of restricted securities as a proxy for establishing a “meme stock” portfolio¹. Furthermore, 7 of these 13 identified stocks by the authors, including Gamestop, underwent short squeezes as evidenced by massive price surges, an increase in available quantity, a decline in short activity, a decline in lendable quality, a decline in average fees for stock lending purposes as well as an overall migration from straight equities to options (2021).

Aloosh et al. (2022), in attempting to measure the contribution of meme stocks to market efficiency, take an almost identical approach but construct two different equally-weighted portfolios – One consisting of 50 stocks which all faced trading restrictions from January 28th, 2021, and the other consisting of 8 stocks which still faced trading restrictions on February 1st, 2021. The authors argue that while retail investors involved in the trading of meme stocks certainly make trades in other securities, the inside information held by Robinhood and other brokerage firms should give good insight as to the securities which are in particular danger of being destabilized and therefore a good indicator as to being a meme stock.

¹ Stocks identified as “meme stocks” were: American Airlines, AMC, BlackBerry, Bed, Bath & Beyond, Castor Maritime Inc., Express, GameStop, Koss, Naked Brand Group, Nokia, Sundial Growers, Tootsie Roll, and TriVaGo.

2.2. Sentiment Analysis and Stock Correlation

A considerable amount of research has been conducted in trying to predict stock market behavior using all means possible. The effect of news, for one, has been one of the more thoroughly studied potential predictors regarding equities movements. Market participants have been interested in the impact of headlines on stock market movements far before personal computers or the internet became popular. In his influential paper, Niederhoffer establishes that world events do indeed have a noticeable impact on movements in the stock markets (1971).

With regards to more recent developments, one of the more renowned and often-cited pieces of research on the topic was published by Bollen, Mao, and Zeng (2010) in which the authors gathered as well as analyzed the collective public mood states (Calm, Alert, Sure, Vital, Kind or Happy) and measured their correlation with the Dow Jones Industrial Average (DJIA). This was achieved by gathering tweets from February 28th, 2008, to December 19th, 2008, normalizing these, and finally applying OpinionFinder as well as Profile of Mood States (POMS), a psychometric instrument with significant precedence in literature. The results of the study, which makes use of their mood state time-series data in conjunction with a self-organizing fuzzy neural network, show an astounding 87% accuracy in predicting the changes in the DJIA closing values. This paper presents an interesting and promising starting point for further discussion but fails to discuss the correlation of individual securities prices with their corresponding sentiment. The study can be seen as an investigation of how the overall mood across a broad population influences movements in price across a broad index of stocks, not however how sentiment of one stock can influence its price.

Taking a similar approach, yet attempting to further narrow their scope, Bing, Chan, and Ou (2014) propose a method by which they can predict price movements not only related to the

DJIA but to specific companies within the NASDAQ. Interestingly, the authors collect twitter data relating to the companies of their choice using keywords relating to that company's products. While this should be able to capture any individual's sentiment towards a company more adequately, it does not directly capture the sentiment of individuals expressing their opinion of a given company with specific regard to its stock price. Using their proposed keyword classification algorithm as well as a support vector machine (SVM), Naïve Bayes, and C.45 classification methods, the authors attempt to classify stock movements into 5 distinct categories depending on the magnitude of stock price change. Overall, they manage to achieve an average prediction accuracy of up to c. 63% across their analyzed industries and all classifiers.

While previous studies have focused largely on unfiltered textual data (especially as sourced on Twitter), Dickinson and Hu (2015) take a different approach by classifying tweets holding little sentiment information as neutral by measuring a tweet's subjectivity/objectivity using Stanford's deep neural network-based NLP processor. Tweets containing advertisements, or other non-user-generated texts were therefore immediately downgraded in terms of their predictive utility. Furthermore, the authors use a training dataset of tweets which they manually trained, to guarantee maximum sentiment prediction accuracy. Tweets were analyzed in terms of correlation with 5-minute stock change intervals. In using popular text processing and NLP algorithms such as n-gram and word2vec, the study manages to achieve a Pearson correlation of up to c. 0.85 for e.g., consumer-facing companies such as Walmart and Microsoft while at the same time achieving a correlation of c. -0.9 for e.g., Goldman Sachs and Cisco Systems. The study is particularly interesting in its manual labeling of a specifically generated dataset because datasets usually used to train NLP models are often not related to the field where their predictions are finally applied. This benefit however comes at the expense of scale, with only

c. 1,000 tweets manually labeled. Furthermore, the study fails to address the class imbalance posed by the fact that most of the tweets in the dataset were neutral with only a small portion being negative or positive, which almost certainly distorts their results and inflates the measured correlation.

3. Methodology

3.1. Data Mining and Extraction

The purpose of this paper is to establish a link between social media sentiment and stock price movement as well as to document a possible “meme effect” in which the sentiment of meme stocks correlates more closely to their underlying security than the sentiment of blue-chip stocks. As discussed earlier, the determinability of this effect is fundamentally dependent on the definition and composition of a meme stock portfolio. For this thesis, Aloosh et al.’s MS8 portfolio will be used as the most indicative selection of stocks (2022) with the exclusion of NAKD due to a merger with Cenntro Automatic Group and resulting price distortions/delisting. This portfolio contains stocks facing the longest trading restrictions by major brokerages and consists of AMC, BB, EXPR, GME, GNUS, KOSS, and NOK. Furthermore, as determined by Allen et al., of these 7 stocks AMC, EXPR, and GME experienced a short squeeze, while the rest did not (2021). This allows for the potential examination of sub-portfolios within the “meme stock” portfolio and the measurement of any potential effects that may arise between sentiment and any security being short squeezed. By the same token, a representative blue-chip portfolio was selected, and the choice was made to use the top ten most heavily weighted constituents of the NASDAQ index excluding GOOG, a separate Alphabet Inc. share class without voting rights that would not offer any additional explanatory value to the research question. The corresponding tickers are AAPL, MSFT, AMZN, TSLA, GOOGL, NVDA, PEP, COST, META, and AVGO.

To answer the research question, this thesis incorporates both stock price data as well as sentiment data. Stock price information extraction was executed using Yahoo Finance’s python module named `yfinance`. Regarding the data on sentiment, the choice was made to extract data only from the social media platform Twitter, both for simplicity and scale’s sake as well as due to the platform’s massive popularity when it comes to exchanging information rapidly especially relating to equities information. A package named `Tweepy` was used to connect to Twitter’s API v2 using Academic-level developer credentials. A python script was devised that mined data related to the chosen securities by making requests for a maximum of 400 tweets (chosen due to API rate limits and throttling issues) every day in the timeframe between 1. December 2019 and 2. October 2022. The date range guarantees that the maximum macroeconomic and idiosyncratic movement is captured, ensuring the inclusion of both the effects of the COVID-19 pandemic as well as the Ukraine war and the following energy crisis. The queries used in the script were the securities’ respective “cashtag” as well as their hashtag connected by a logical OR (e.g., “\$GME OR #GME”). In the case of Facebook, both the former cashtag “\$FB” and the current cashtag “\$META” were used to reflect the company’s name change. The construction of these queries ensures that the majority of tweets returned by the scraper are related to the company’s listed security. It cannot, however, be excluded that non-related tweets are collected. Finally, a filter was applied to return only English non-retweet tweets. The following table gives an overview of the number of tweets downloaded for each security (incl. duplicates, which were later removed):

Table 1: Total tweets downloaded for each meme stock

AMC	BB	EXPR	GME	GNUS	KOSS	NOK
256,636	308,692	40,954	104,199	102,566	39,424	111,384

Table 2: Total tweets downloaded for each NASDAQ stock

AAPL	AMZN	AVGO	COST	GOOGL	META	MSFT	NVDA	TSLA	PEP
329,084	330,319	70,746	84,912	164,517	211,375	227,482	234,182	213,975	106,248

3.2. Data Cleaning and Preparation

Before any textual adjustments were made to the tweets, all duplicate tweets were dropped from the dataset in an attempt to remove spam tweets as well as potential advertisements, which bear no significant sentiment information. Even excluding the muddying consequences of spam, sentiment analysis for social media texts of any kind is notoriously difficult to execute due to a variety of factors including sarcasm, context-based semantics, emojis as well as hyperlinks, and user mentions. This noise holds no relevant information to the models at hand while obfuscating potential true sentiment (Haddi, Liu, and Shi 2013). It is therefore of paramount importance that the text in question (in this case hundreds of thousands of tweets) be adequately cleaned and converted into a textual representation that natural language processing (NLP) algorithms can effectively analyze. In the case of the tweets, this cleaning process consisted of several steps. First, the tweets were all converted to lowercase letters. Then, regular expressions were used to remove all username mentions (e.g., “@elonmusk”) as well as hyperlinks. Next, all punctuation within the tweets was removed, retaining only regular characters and whitespaces. Following this step, traditionally stop words such as “is”, “not”, “won’t”, and “to” are removed, which for most NLP tasks do not add any predictive value to the overall sentence but can slow down an NLP model (Sarica and Luo 2021). It has been widely shown that most NLP model evaluation metrics are considerably benefitted from the removal of stop words (Haddi, Liu, and Shi 2013). Within the scope of sentiment analysis, however, stop words have an undeniable importance. Sentence negation provided by the stop word “don’t” makes the semantic difference between “I like the stock” and “I don’t like the stock”. While it is true that model

vector spaces and vocabulary sizes decrease with stop word removal, sentiment classification accuracy demonstrably suffers (Jianqiang and Xiaolin 2017). The choice was therefore made not to remove any negating stop words from the corpus of tweets. In the last step of the cleaning process, the remaining sentence was normalized, using lemmatization, a commonly used technique aiming to establish a word form’s most basic root to further simplify the text and to detract further unnecessary information. As an example of the total text transformation, consider the following tweet:

“#GameStop up 5.1% = \$28.5 million gain. \$GME has made notable gains in 11 of the last 38 trading sessions. <https://t.co/1LYRF6E5Pb> #stockMarket #stocks #trading #market #investing #NASDAQ #NYSE #consumerDurables #computer #electronics <https://t.co/7brgJkEW6P>”

After preparation and cleaning, the tweet is transformed into the following:

“up million gain made notable gain last trading session”

As can be observed, the cleaning and preparation process considerably shortens and simplifies relevant tweets, allowing for NLP models to run with much greater efficiency and accuracy.

Consider, summarily, a chart of the entire cleaning process:

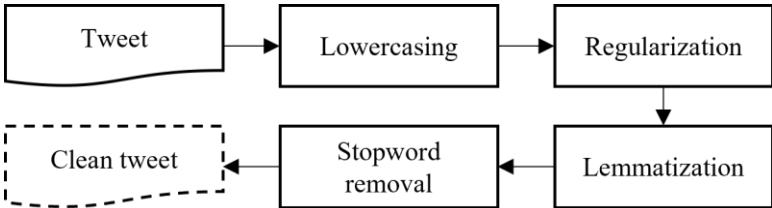


Figure 1: Tweet cleaning process

3.3. Model Selection and Preparation

3.3.1. Textblob

Textblob is a pre-trained module built to be functional on top of Python 2 and 3 architectures (Loria S 2020). While the library features a multitude of useful NLP tools such as spelling

correction, classification, and noun phrase extraction, it was used in the context of this thesis solely for its sentiment analysis functionality. Like many similar sentiment classifiers, Textblob was trained using a dataset of pre-classified movie reviews. Upon presenting the classifier with an unknown text, it makes use of a probability-based Naïve Bayes classifier to calculate the text's polarity and subjectivity by employing a lexicon of pre-defined words and their corresponding polarity. To answer the research question, it was decided that only the polarity result of Textblob's sentiment analyzer was relevant.

3.3.2. VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) is another sentiment analysis library built for Python but ported to a multitude of other programming languages. Much like Textblob, VADER employs a rule-based lexicon approach to sentiment analysis specifically attuned to the analysis of sentiment on social media using a multitude of different training sets including Amazon review snippets, tweets, and movie reviews (Hutto and Gilbert 2014). The classifier calculates a compound score based on the valence scores of each word within a given text while applying further pre-defined rules relating to e.g., word-order sensitivity or degree modifiers. In the scope of this thesis, the compound score was therefore used to represent overall tweet sentiment for every security's tweets. What is particularly interesting about VADER is the ability to refine its predetermined scoring dictionary with custom words. Most pre-trained NLP models (including Textblob) have difficulties when it comes to specific language and slang used in certain sectors and communities, for example, unable to determine whether the term "bullish" is positive or negative. VADER's default dictionary was therefore adjusted to include a custom dictionary of meme stock-related jargon. This allowed the model to achieve significantly improved accuracy.

3.3.3. Gated Recurrent Unit

An additional method used in the context of this thesis to classify a tweet's sentiment was to construct and manually train a deep recurrent neural network using the gated recurrent unit (GRU) architecture. The approaches discussed previously employ a lexicon-based approach to digesting language information, which has difficulty inferring context from several words in sequence, which RNNs can achieve through their recursive nature and ability to therefore extract sequentially conveyed information. To circumvent this problem, the Long Short-Term Memory (LSTM) network was conceived, which introduces several so-called "gates" within each neuron (referred to as cells), to control the flow of information between layers (Hochreiter and Schmidhuber 1997). Recently, the design of this LSTM architecture has been innovated in form of the GRU network, which is conceptually and mathematically simpler (Cho et al. 2014). The exact mathematical derivation of such models exceeds the scope of this paper, but research has shown that GRU networks are viable alternatives to LSTM models and valuable in NLP tasks and sequence modeling (Chung et al. 2014). When constructing any neural network, it is of utmost importance to provide it with a relevant and labeled dataset in as much granularity as possible. Furthermore, in the best case, this dataset would contain tweets related to or directly analogous to the tweets for which the trained model would make predictions later. In the case of Twitter data, procuring such a dataset presented considerable difficulty. Therefore, to generate as much relevant training data as possible, several datasets were concatenated into one which included a dataset containing sentiment relating to various entities commonly discussed on Twitter, a dataset relating to the 2020 U.S. GOP debate as well as a shorter dataset containing tweets relating to Apple's stock. The total dataset used in training and testing amounted to 93,178 tweets. The tweets were each labeled either positive, negative, or neutral. The tweets were cleaned and prepared using the procedure stated before, while additionally dropping any tweets below 3 words, as these were rather unlikely to hold much predictive power and

demonstrably decreased model performance during training. Tweets were then tokenized, padded, and vectorized to create data that can be more easily understood by the model. This vector is then encoded into a dense representation matrix, where similar words have similar encoding which is learned by the network over its iterations. The neural network architecture was constructed as follows: First, the tokenized tweets were encoded into a dense representation matrix using an embedding layer. After the embedding layer, the model features a GRU layer with 128 cells and a dropout regularizer of 0.1 (meaning 10% of all weights are randomly dropped in each iteration to avoid overfitting). The outputs of this layer are then passed to a batch normalization layer which normalizes inputs to the next layer and significantly speeds up the overall training procedure. The normalization is followed by another GRU layer with 128 cells, which is followed by a dense output layer with 3 neurons which represent our 3-class classification problem.

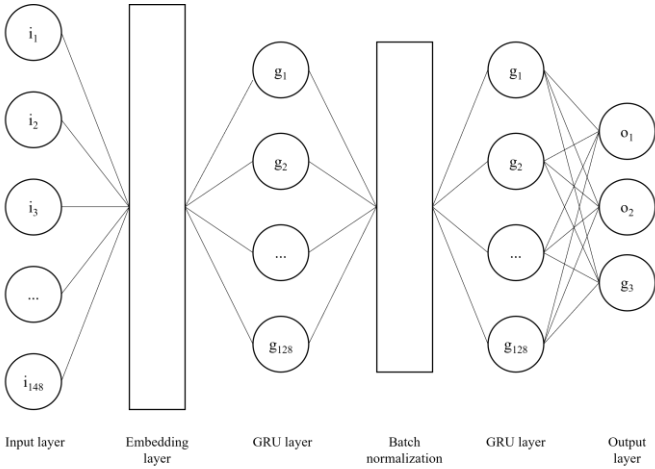


Figure 2: Gated recurrent neural network architecture

The model makes use of a sparse categorical cross-entropy loss function with an Adam optimizer and an initial learning rate of 0.05. Furthermore, the train data was rebalanced to avoid misclassification. The model was trained for 28 epochs with a test loss of 0.6585 and a test accuracy of 80.95%. Overall, the model does seem to have a slight overfitting problem, but with a test accuracy of over 80%, this can be disregarded as having minimal impact.

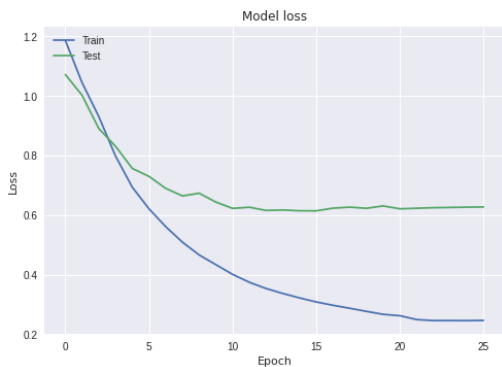


Figure 4: Model training loss

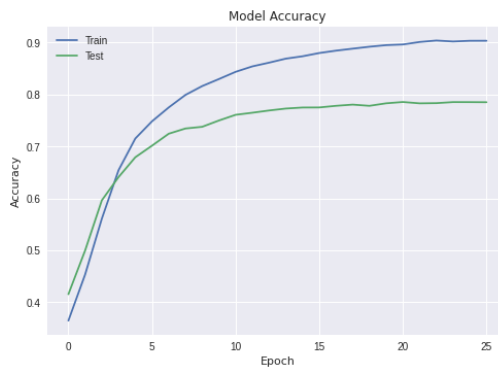


Figure 3: Model training accuracy

3.3.4. roBERTa

The final, and perhaps most powerful tool used to classify tweet sentiment is a successor of the bidirectional encoder representations from transformers (BERT) developed at Google AI by Jacob Devlin and his team (Devlin et al. 2018). BERT is a pre-trained general language representation that essentially learns to understand language and context in an unsupervised manner by simultaneously solving masked language modeling (MLM) and next sentence prediction (NSP). As no labels are needed due to the nature of these tasks (since the correct labels already exist in any given text), massive bodies of texts can be used to train BERT such as BooksCorpus with 800m words and the English Wikipedia with 2,500m words. After such a model has been trained, it can then be fine-tuned for many kinds of NLP tasks such as question and answer problems, named entity recognition, or sentiment classification by providing a state-of-the-art word embedding space as discussed in section 3.3.3. The advantage over LSTM / GRU networks is the ability to parse information and train using parallelization while recurrent neural networks must process information sequentially, thereby reducing efficiency. roBERTa (robust BERT approach) is an implementation of BERT with slight alterations to drive significant performance gains while using a more recent training dataset and was therefore the choice for classifying sentiment in the scope of this paper (Liu et al. 2019).

3.4. Merging Classifiers and Preparing Correlational Analysis

After having prepared and applied the classifiers, a weighted average was applied to unify the separate predictions into one and to operationalize the “wisdom of the crowd” concept. All classifiers’ predictions were manually inspected for a few randomized tweets. Consider a table of some of the tweets and their associated predictions.

Table 3: Predicted tweet sentiment by classifier

Tweet	Textblob	VADER	GRU	roBERTa
gamestops stock look like value trap read	0.00	0.38	Negative	Negative
call buyer super active today	0.10	0.77	Positive	Positive
discount still up grab get em	0.00	0.00	Positive	Neutral
wow amazing gain today excellent work join discord today	0.57	0.94	Positive	Positive
silent over month bullish af alex	0.00	0.61	Neutral	Neutral
bear market wont end lower level	0.00	-0.48	Positive	Negative
well said buy dip otherwise	0.00	0.73	Positive	Neutral
fed crash market via	0.00	-0.61	Neutral	Neutral
many canceled misguided know let destroy economy	0.15	-0.54	Negative	Negative

As becomes quite clear, the tweets are difficult to interpret even as a human, and the models face even greater difficulties. Generally, it seems that Textblob has the most trouble predicting accurate sentiment while both VADER and the neural network models agree most of the time, confirmable also via inter-classifier sentiment correlation. VADER, as well as the neural network models, will therefore have a higher weighting in the construction of the overall sentiment evaluation.

The information within the datasets containing the financial data had a daily frequency due to yfinance’s limitations, therefore the prediction datasets had to be grouped by day and the

combined sentiment of each day was averaged. The use of the average negates the distorting effects of any certain security having more tweets than another. Following this step, a predictive analysis could be performed.

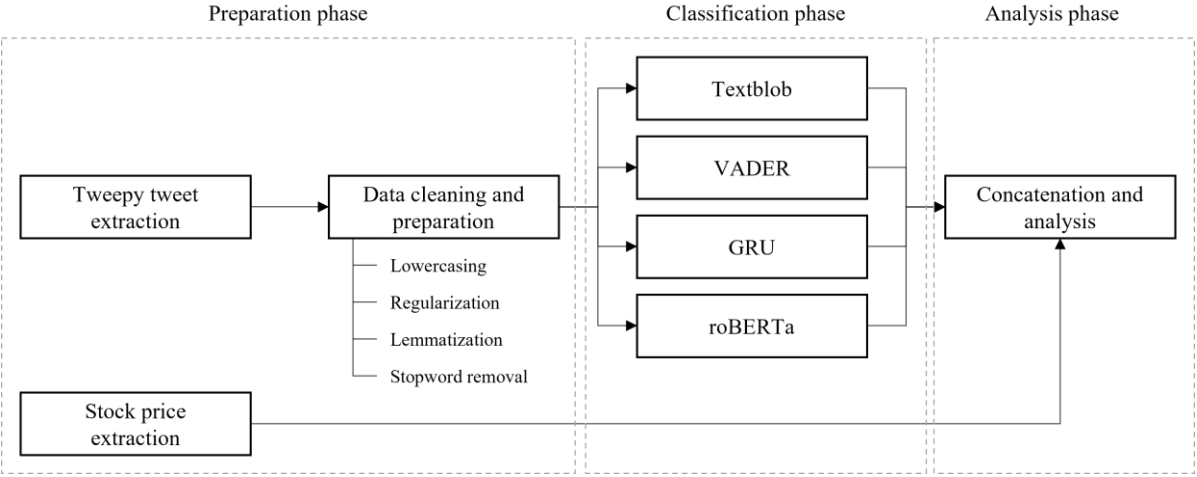


Figure 5: Overall process & methodology

4. Evaluation

4.1. Time-lagged Cross-correlation

To answer the research question at hand, the objective was to establish not only the existence of a correlation between daily sentiment and daily closing price of every respective security within the meme stock portfolio but to also determine whether this correlation differed significantly compared to the portfolio of blue-chip stocks. As the datasets consist of time-series data they must be subject to rigorous examination and transformation before statistically relevant conclusions can be drawn. In particular, correlational analysis can be heavily skewed by underlying linear (or non-linear) trends in the datasets which can be caused due to autocorrelation within each time series for example. In the case at hand, for example, the sentiment of σ_t is most likely slightly correlated with the sentiment of σ_{t-1} . The same conclusion could hold for stock price developments, as was rigorously documented in finance in the form of the momentum effect (Jegadeesh and Titman 1993). As becomes evident from the following

example graphs of the standard-scaled weighted average sentiment plotted against the standard-scaled closing price, a linear trend in some of the datasets' time-series indeed exists.

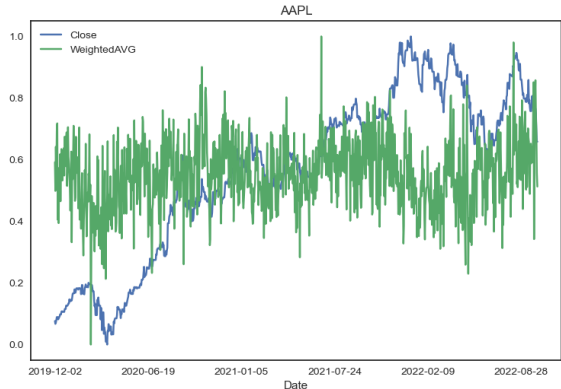


Figure 6: AAPL price and sentiment

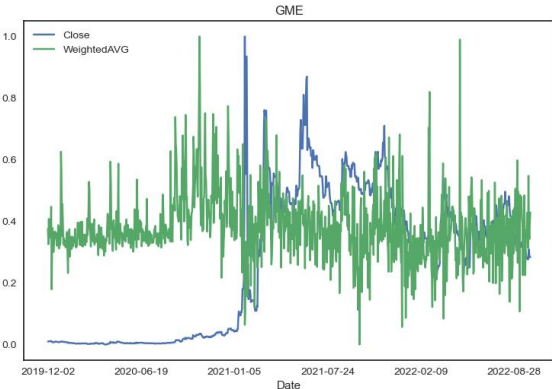


Figure 7: GME price and sentiment

To ensure statistical reliability, time-series must be stationary and detrended. To check for this, the commonly used augmented Dicky-Fuller test (ADF) was used, which tests for the presence of a unit root within the time series data (Cheung and Lai 1995). After stationarity was ensured using first-order differencing for every separate dataset, time-lagged cross-correlation for various degrees of lag was calculated. The intuition behind this is the fact that it could establish a sort of directionality within the correlation coefficients (e.g. an increase in the sentiment of observation o_{n-1} leads to an increase in the stock price for observation o_n). For the correlation coefficients themselves, Spearman's r was used, as there was no immediate reason to assume a strictly linear relationship between sentiment and stock price mandating the use of Pearson's r . After calculation, it becomes apparent that the highest spearman correlation for most securities results from a delay of 0 days, apart from GNUS and KOSS, with a maximum correlation when their price trails sentiment by 3 days. From this analysis, it is therefore not immediately possible to infer any sort of directionality in terms of information transfer between sentiment and close. The following figure visualizes the analysis for the example of GME:

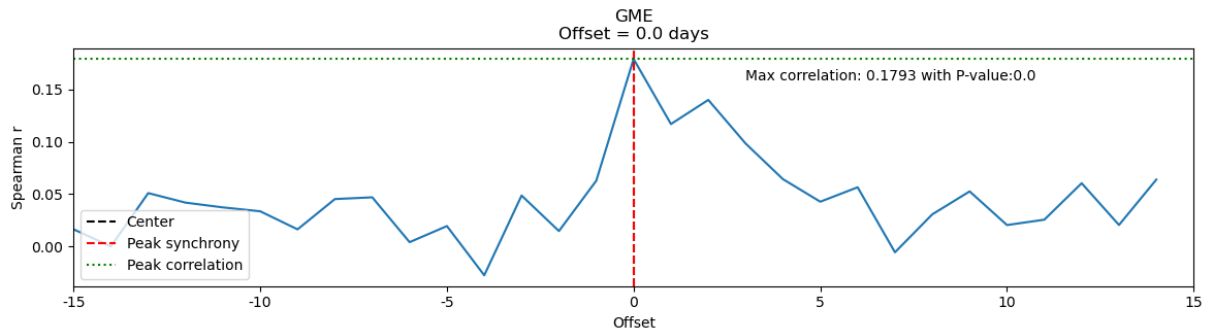


Figure 8: Lagged cross-correlation for GME

To ensure maximum comparability, 7 of the 10 NASDAQ stocks were picked at random.

Table 4: Correlation and p-values for meme stocks

	AMC	BB	EXPR	GME	GNUS	KOSS	NOK	Avg.
Spearman r	0.275	0.098	0.160	0.179	0.133	0.088	0.093	0.147
p-value	<0.01	<0.01	<0.01	<0.01	<0.01	0.019	0.013	-

Table 5: Correlation and p-values for NASDAQ stocks

	AAPL	AMZN	AVGO	COST	META	NVDA	TSLA	Avg.
Spearman r	0.409	0.332	0.138	0.224	0.217	0.428	0.318	0.295
p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	-

For the meme stocks, arguably, significant irregularities in trading occurred especially on the retail front, as Robinhood and other brokers heavily restricted or even banned trading in these stocks in the period from the 28th of January 2021 to the 5th of February 2021. Therefore, the analysis was repeated, excluding this range from the datasets.

Table 6: Correlation and p-values for meme stocks excl. restriction period

Security	AMC	BB	EXPR	GME	GNUS	KOSS	NOK	Avg.
Correlation	0.273	0.091	0.155	0.174	0.149	0.081	0.090	0.145
p-value	<0.01	0.015	<0.01	<0.01	<0.01	0.032	0.017	-

Table 7: Correlation and p-values for NASDAQ stocks excl. restriction period

Security	AAPL	AMZN	AVGO	COST	META	NVDA	TSLA	Avg.
Correlation	0.403	0.336	0.145	0.222	0.212	0.431	0.320	0.296
p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	-

Evidently, the correlation does not increase when the restriction period is excluded from the overall sample period. This may seem counterintuitive at first, but it is important to remember

that the restrictions were mainly faced by retail investors while institutional investors were still able to make trades potentially informed by sentiment with massive volumes. It was therefore decided to retain the period facing restrictions in all further analyses. Finally, for each of the stocks, a Granger causality analysis was performed to establish a degree of predictive certainty in terms of signal information flow between two time-series (Granger 1969). The analysis tests for the null hypothesis that series x does not cause series y (and vice versa). The results are shown below:

Table 8: Granger causality for meme stocks

Security	AMC	BB	EXPR	GME	GNUS	KOSS	NOK
p-value (Sentiment precedes close)	<0.05*	<0.01**	>0.05	>0.05	>0.05	>0.05	>0.05
p-value (Close precedes sentiment)	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05

Note: ** - statistically significant at .01 level, * - statistically significant at .05 level

Table 9: Granger causality for NASDAQ stocks

Security	AAPL	AMZN	AVGO	COST	META	NVDA	TSLA
p-value (Sentiment precedes close)	<0.05*	<0.01**	>0.05	<0.01**	<0.01**	>0.05	<0.01**
p-value (Close precedes sentiment)	<0.01**	>0.05	>0.05	<0.01**	>0.05	>0.05	>0.05

Note: ** - statistically significant at .01 level, * - statistically significant at .05 level

For all meme stocks except AMC and BB, there is no readily discernible directionality to the information flow between sentiment and closing price, however, for the NASDAQ stocks, it appears as though sentiment tends to be more relevant in explaining moves in the closing price as compared to vice versa as measured by the larger statistical significance in the first row.

4.2. Binary Classification

Additional predictive analysis in the form of binary classification was executed by converting each dataset's closing price into a binary variable. The variable was coded as 1 if a stock's price increased in o_n compared to o_{n-1} and 0 if the opposite was true. 3-day off-center rolling weighted average of the sentiment was used as the sole explanatory variable. Several binary classifiers were trained, evaluated using 10-fold cross-validation, and then compared using classification

accuracy as a metric. Classifiers used were Logistic Regression (LR), Linear Discriminant Analysis (LDA), k-nearest Neighbors (KNN), Decision Tree (CART), Gaussian Naïve Bayes (NB) as well as a Support Vector Machine (SVM) with default parametrization (i.e. no hyperparameter optimization). Once again, the portfolio of meme stocks was compared to the same selection of NASDAQ stocks. The results are as follows:

Table 10: Binary classification accuracy for meme stocks

Security/Classifier	LR	LDA	KNN	CART	NB	SVM	Avg.	Max
AMC	59.7%	59.4%	56.5%	54.4%	60.3%	59.1%	58.2%	60.3%
BB	55.9%	55.9%	50.3%	45.9%	54.4%	51.5%	52.3%	55.9%
EXPR	57.4%	57.4%	53.8%	55.3%	58.5%	58.5%	56.8%	58.5%
GME	50.0%	50.0%	50.3%	51.5%	55.0%	56.2%	52.2%	56.2%
GNUS	49.7%	49.7%	52.1%	50.0%	50.9%	45.6%	49.7%	52.1%
KOSS	57.9%	57.9%	54.4%	45.0%	57.4%	55.9%	54.8%	57.9%
NOK	50.9%	50.9%	49.7%	49.1%	51.2%	46.8%	48.8%	50.9%
Total Avg.							53.3%	56.0%

Table 11: Binary classification accuracy for NASDAQ stocks

Security/Classifier	LR	LDA	KNN	CART	NB	SVM	Avg.	Max
AAPL	59.7%	59.7%	55.0%	51.2%	58.8%	58.8%	57.2%	59.7%
AMZN	54.7%	54.4%	55.3%	52.9%	55.6%	57.6%	55.1%	57.6%
AVGO	52.6%	52.6%	46.2%	45.9%	51.5%	52.6%	50.2%	52.6%
COST	56.2%	56.2%	51.8%	53.5%	58.8%	55.6%	55.3%	58.5%
META	53.8%	53.8%	53.2%	54.4%	53.5%	49.7%	53.1%	54.5%
NVDA	56.8%	56.8%	51.5%	45.0%	56.8%	60.0%	54.5%	60.0%
TSLA	58.5%	58.5%	52.9%	53.5%	58.2%	58.2%	56.7%	58.5%
Total Avg.							54.6%	57.3%

Using the last three days' weighted average Twitter sentiment, in the case of meme stocks, blended classifier predictions feature an accuracy of 53.3% on average. For NASDAQ stocks, this accuracy is increased to c. 54.6%. The maximum accuracy predictors for the meme stocks average 56.0% compared to 57.3% for the NASDAQ stocks. Due to the stochastic nature of the classification, this can vary slightly every time the code is run.

4.3. Hypothetical Portfolio Strategy and Back Testing

While the aim of this research paper is by no means to provide an actionable trading strategy

based on sentiment classification, the analysis is highly suggestive of such a strategy potentially creating value for an investor. To test this, a simple algorithm was conceived and back tested using `backtest.py`, a common framework for testing trading strategies. The strategy was restricted to long positions only and invested all available equity with every trade. This implies that only one position was active simultaneously. The signal to buy was given when 3-day rolling weighted average sentiment crossed a threshold of 0.5, and the signal to close the position was given when that same number crossed a threshold of -0.5. A broker commission of 0.002 was assumed and starting equity was 10,000 USD with a 95% stop-loss and a 200% take-profit. Overall, the strategy generated average returns of c. 30% during the holding period with a Sharpe ratio of 0.29 for a portfolio of equally weighted meme stocks and average returns of c. 85% with a Sharpe ratio of 0.75 for an equally weighted portfolio of the seven NASDAQ stocks. While this result seems largely in line with the preceding results regarding the increased predictability of the blue-chip portfolio compared to the meme stocks, the parameters of the strategy as well as its sentiment threshold could be tweaked in future research to generate vastly different results.

4.4. Discussion

There exists an undeniable correlation between a security's Twitter sentiment and its closing price. This is supported not only by the time-lagged correlation results but also by the Granger causality analysis and the non-trivial prediction accuracy achieved over a variety of binary classifiers. What is additionally of great interest is the fact that within the meme stock portfolio, the three securities facing a short squeeze (AMC, EXPR, GME) were also the three securities with the highest price correlation to sentiment. Apart from KOSS, these three securities also featured the highest accuracy in the binary classification problem. Therefore, it seems reasonable to assume that short squeezes have a considerable positive effect in their

contribution to the correlation between sentiment and price. Intuitively, this makes sense as stocks being short squeezed attract large social media attention and investor interest, with more sensibilization of the price to subsequent moves in social media sentiment. In contrast to initial intuition regarding the difference in predictability between meme stocks and blue-chip stocks, however, the price of the former seems to correlate less significantly with sentiment than that of the latter, across virtually all analyses conducted in this thesis. The portfolio of blue-chip stocks has a larger Spearman correlation, more significant p-values for the Granger causality analysis, higher binary classification accuracy as well as better performance in a back tested trading scenario. This result is especially counterintuitive given the fact that a large portion of the research on sentiment affecting stock prices seems to have been initiated in response to the meme stock phenomenon in the first place. There are several possibilities as to why this is the case, the most likely of them having to do with the accuracy of sentiment classification. The tweets relating to meme stocks face a far higher rate of spam, unrelated content or misinformation, bot activity, and a general deterioration in intelligible language due to being so wildly popular and garnering an almost cult-like following with highly domain-specific jargon. This makes it much more difficult for the sentiment classification algorithms to accurately classify sentiment for the meme stocks leading to a divergence in co-movement of predicted sentiment and price. Another very likely reason is the fact that many users may have preferred to share information on potential trades on Reddit as opposed to on Twitter, which is certainly not the case for the blue-chip portfolio as these stocks are not the type to be vigorously discussed on “/r/wallstreetbets”.

5. Limitations

This thesis faces some limitations, especially concerning the gathering of Twitter data as well as the sentiment classification portion of the analysis. First, regarding the gathering of data on

Twitter, the language filter may not always be accurate, and not all tweets gathered were English. Furthermore, the data gathered may not be relevant whatsoever, and despite best efforts of removing spam, sexually explicit content, as well as advertisements, a completely clean dataset could not be guaranteed. There was also no efficiently scalable process to remove bots posting the same tweet repeatedly, even after removing duplicates using fuzzy matching. An additional limitation of the data-gathering process of this paper is the rate limit set by Twitter's API, imposing a cap on the number of tweets a day. Relating to the sentiment analysis, there is one glaring limitation that has to do with entity relation. The tweet "I love \$GME but think \$AMC is totally overvalued", ("i love but think is totally overvalued" after the cleaning process) contains two sentiments relating to two different securities yet the sentiment classifiers will assign only one compound sentiment which introduces significant inaccuracy.

6. Future Research

As the flow of information relating to equities and other traded securities increasingly moves to the internet, the topic of sentiment analysis and classification presents a striking opportunity for future research. Particularly in the field of cryptocurrency, which is a market that is arguably driven almost entirely by emotional responses and irrational trading behavior with highly volatile assets, sentiment analysis could be extremely relevant. Additionally, while this study made use of c. 3 years of data, future studies could more accurately capture co-movements between sentiment and price movements by broadening the time frame significantly. To combat the entity relation limitation discussed above, further research could add named entity recognition step into the methodology. Finally, future research could focus on aggregating data sources by making use of not only Twitter, Stocktwits, or Wallstreetbets alone, but combining all of these into a vast set of social media data relating to securities and perhaps cryptocurrencies. All these points could be valuable opportunities in terms of future research.

7. Conclusion

This paper has attempted to establish a relationship between sentiment and stock price movements, by making use of a blended sentiment classification approach combined with binary stock movement classification methodology. The findings allow us to posit an answer to the research question: *Is the movement of meme stock prices predictable using social media sentiment and how does the predictability differ from traditional blue-chip stocks?*

Overall, there is indeed some degree of stock price predictability using the suggested blended sentiment classifier approach. This correlation exists for both meme stocks – where stocks affected by a short squeeze had the highest correlation – and blue-chip stocks, but the magnitude was more significant for the blue-chip stocks. Furthermore, for both meme stocks and the NASDAQ portfolio, there is some degree of discernable directionality to the relation between sentiment and stock price close as documented by the Granger causality analysis which is larger for the NASDAQ portfolio and generally implies that price lags sentiment. Finally, making use of binary classifiers without hyperparameter optimization results in average classification accuracy for the daily stock price movement of c. 55% for the NASDAQ portfolio compared to c. 53% percent for the meme stock portfolio. A cursory analysis using a sentiment-derived back tested strategy led to higher returns for the blue-chip portfolio than for the meme stock portfolio, most probably because of the increased sentiment predictability. In answering the research question, this paper has contributed to existing literature by creating a selective overview of the most relevant studies relating to the sentiment classification of stocks on Twitter while relating this specifically to the emergent phenomenon of meme stocks. Additionally, an analysis was conducted relating to the directionality of sentiment information and its effect on stock prices. Finally, a gated recurrent neural network was custom trained specifically for Twitter sentiment classification for meme stocks with considerable testing accuracy.

8. References

- Allen, Franklin, Marlene Haas, Eric Nowak, Matteo Pirovano, Angel Tengulov, Patrick Bolton, François Degeorge, et al. 2021. “Squeezing Shorts Through Social Media Platforms *.” <https://ssrn.com/abstract=3823151>.
- Aloosh, Arash, Hyung-Eun Choi, and Samuel Ouzan. 2022. “The Tail Wagging the Dog: How Do Meme Stocks Affect Market Efficiency?” <https://ssrn.com/abstract=3839832>.
- Bing, Li, Keith C.C. Chan, and Carol Ou. 2014. “Public Sentiment Analysis in Twitter Data for Prediction of a Company’s Stock Price Movements.” In *Proceedings - 11th IEEE International Conference on E-Business Engineering, ICEBE 2014 - Including 10th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC 2014 and 1st Workshop on E-Commerce Engineering, ECE 2014*, 232–39. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICEBE.2014.47>.
- Bollen, Johan, Huina Mao, and Xiao-Jun Zeng. 2010. “Twitter Mood Predicts the Stock Market,” October. <https://doi.org/10.1016/j.jocs.2010.12.007>.
- Cheung, Yin-Wong, and Kon S. Lai. 1995. “Lag Order and Critical Values of the Augmented Dickey–Fuller Test.” *Journal of Business & Economic Statistics* 13 (3): 277–80. <https://doi.org/10.1080/07350015.1995.10524601>.
- Cho, Kyunghyun, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches,” September. <http://arxiv.org/abs/1409.1259>.
- Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” December. <http://arxiv.org/abs/1412.3555>.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding,” October.

<http://arxiv.org/abs/1810.04805>.

- Dickinson, Brian, and Wei Hu. 2015. "Sentiment Analysis of Investor Opinions on Twitter." *Social Networking* 04 (03): 62–71. <https://doi.org/10.4236/sn.2015.43008>.
- Fitzgerald, M. 2021. "Robinhood Restricts Trading in GameStop, Other Names Involved in Frenzy." CNBC. January 28, 2021. <https://www.cnbc.com/2021/01/28/robinhood-interactive-brokers-restrict-trading-in-gamestop-s.html>.
- Godfrey, Keith R.L. 2016. "Detecting the Great Short Squeeze on Volkswagen." *Pacific Basin Finance Journal* 40 (December): 323–34. <https://doi.org/10.1016/j.pacfin.2016.02.001>.
- Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424. <https://doi.org/10.2307/1912791>.
- Haddi, Emma, Xiaohui Liu, and Yong Shi. 2013. "The Role of Text Pre-Processing in Sentiment Analysis." In *Procedia Computer Science*, 17:26–32. Elsevier B.V. <https://doi.org/10.1016/j.procs.2013.05.005>.
- Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hutto, C J, and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." <http://sentic.net/>.
- Jegadeesh, N, and S Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*.
- Jianqiang, Zhao, and Gui Xiaolin. 2017. "Comparison Research on Text Pre-Processing Methods on Twitter Sentiment Analysis." *IEEE Access* 5: 2870–79. <https://doi.org/10.1109/ACCESS.2017.2672677>.
- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. "RoBERTa: A Robustly Optimized BERT Pretraining Approach," July. <http://arxiv.org/abs/1907.11692>.

Loria S. 2020. "Textblob Documentation."

Newman, Nic, Richard Fletcher, Anne Schulz, Simge Andi, Craig T Robertson, and Rasmus

Kleis Nielsen. 2021. "The Reuters Institute Digital News Report 2021."

Niederhoffer, Victor. 1971. "The Analysis of World Events and Stock Prices." *Source: The*

Journal of Business. Vol. 44. <https://www.jstor.org/stable/2351663?seq=1&cid=pdf->.

Sarica, Serhad, and Jianxi Luo. 2021. "Stopwords in Technical Language Processing." *PLoS*

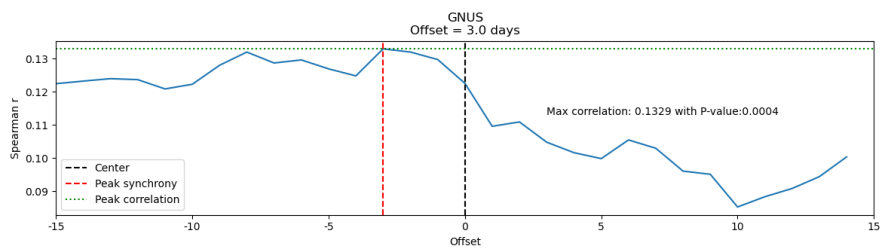
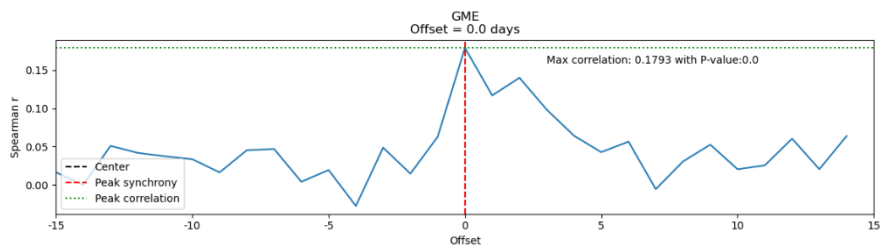
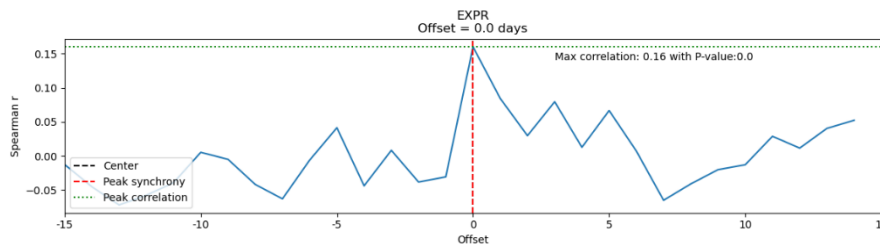
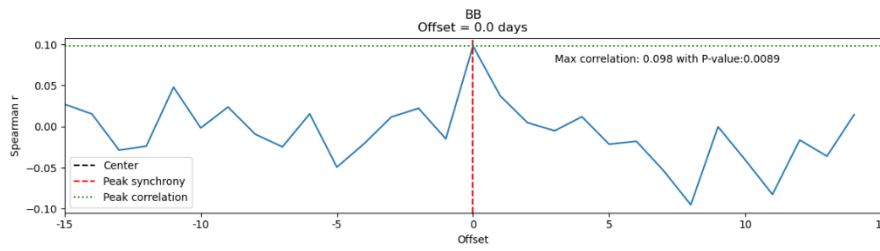
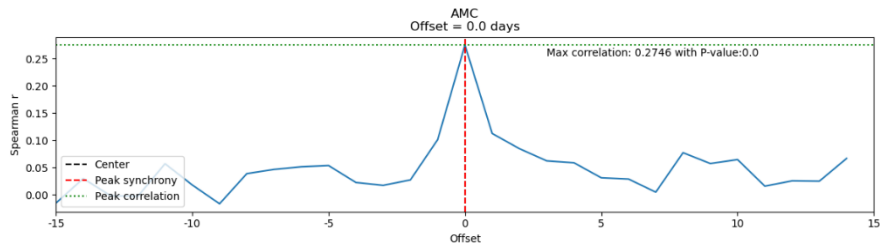
ONE 16 (8 August). <https://doi.org/10.1371/journal.pone.0254937>.

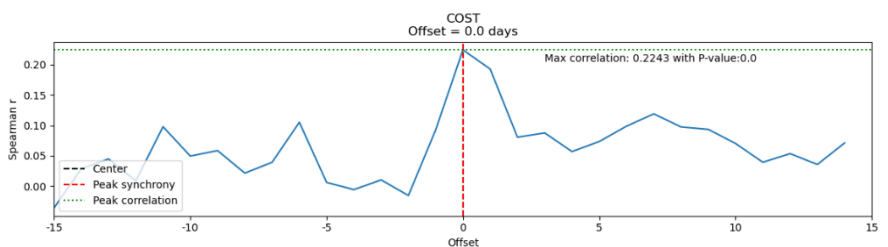
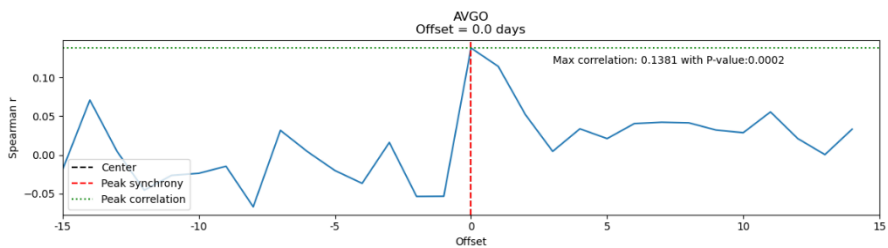
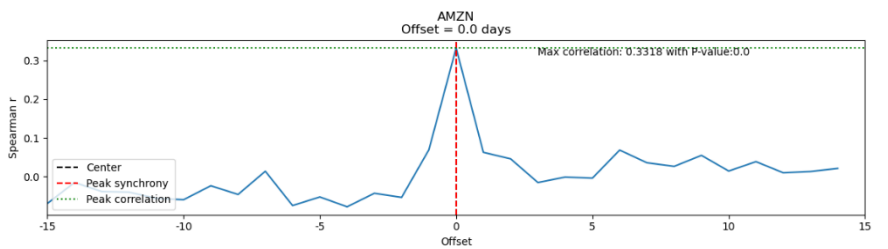
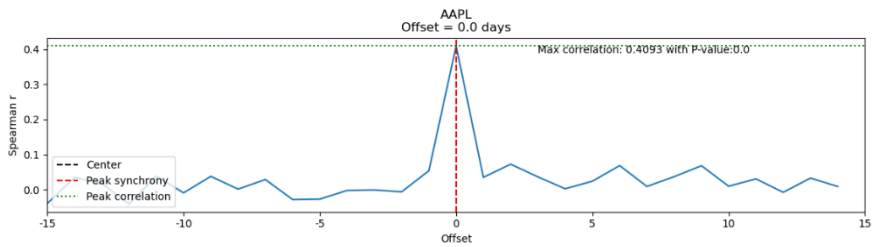
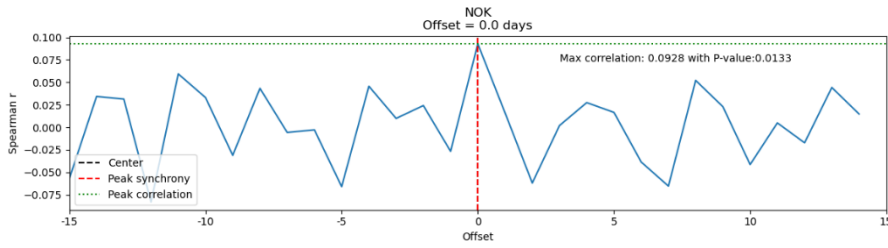
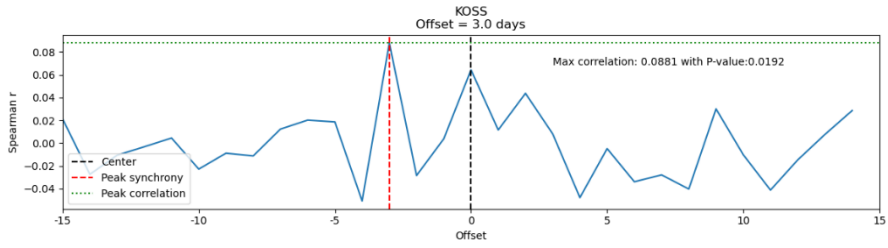
9. List of Figures and Tables

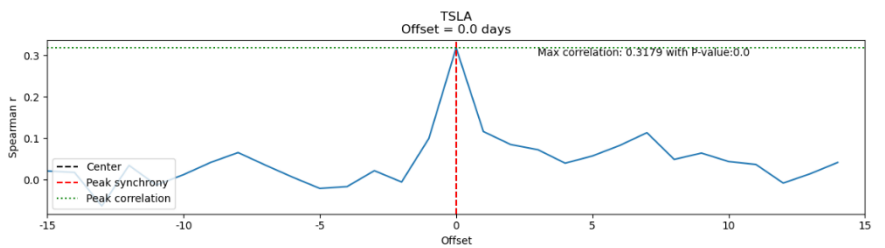
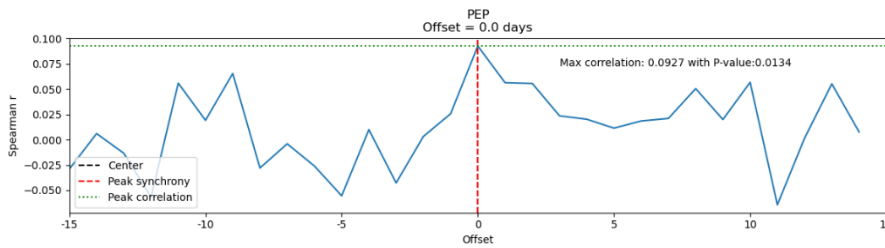
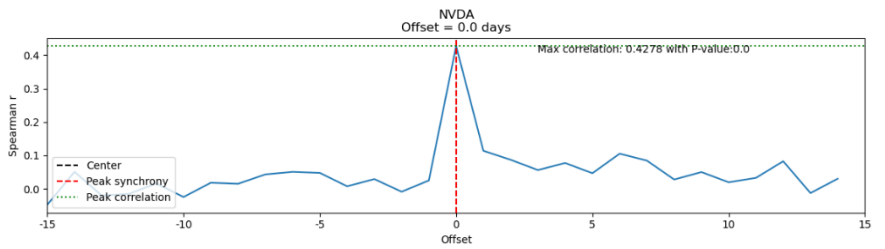
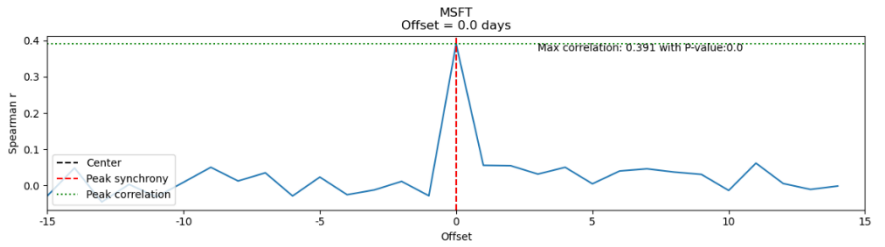
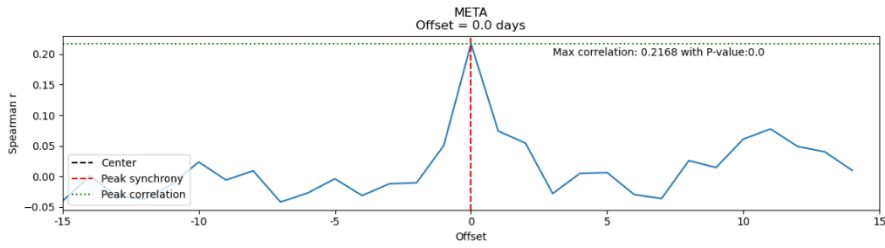
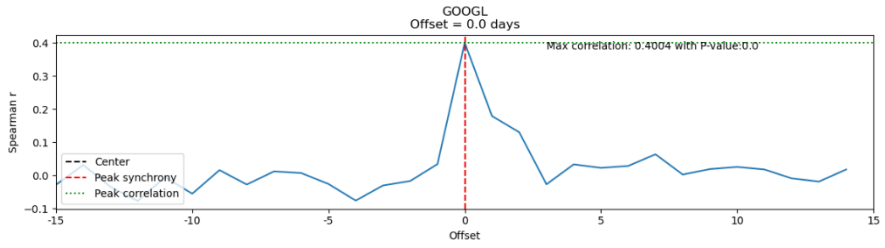
Table 1: Total tweets downloaded for each meme stock	9
Table 2: Total tweets downloaded for each NASDAQ stock	10
Table 3: Predicted tweet sentiment by classifier	16
Table 4: Correlation and p-values for meme stocks.....	19
Table 5: Correlation and p-values for NASDAQ stocks.....	19
Table 6: Correlation and p-values for meme stocks excl. restriction period.....	19
Table 7: Correlation and p-values for NASDAQ stocks excl. restriction period.....	19
Table 8: Granger causality for meme stocks	20
Table 9: Granger causality for NASDAQ stocks	20
Table 10: Binary classification accuracy for meme stocks	21
Table 11: Binary classification accuracy for NASDAQ stocks	21
Figure 1: Tweet cleaning process	11
Figure 2: Gated recurrent neural network architecture	14
Figure 4: Model training accuracy	15
Figure 3: Model training loss	15
Figure 5: Overall process & methodology	17
Figure 6: AAPL price and sentiment.....	18
Figure 7: GME price and sentiment	18
Figure 8: Lagged cross correlation for GME	19

10. Appendices

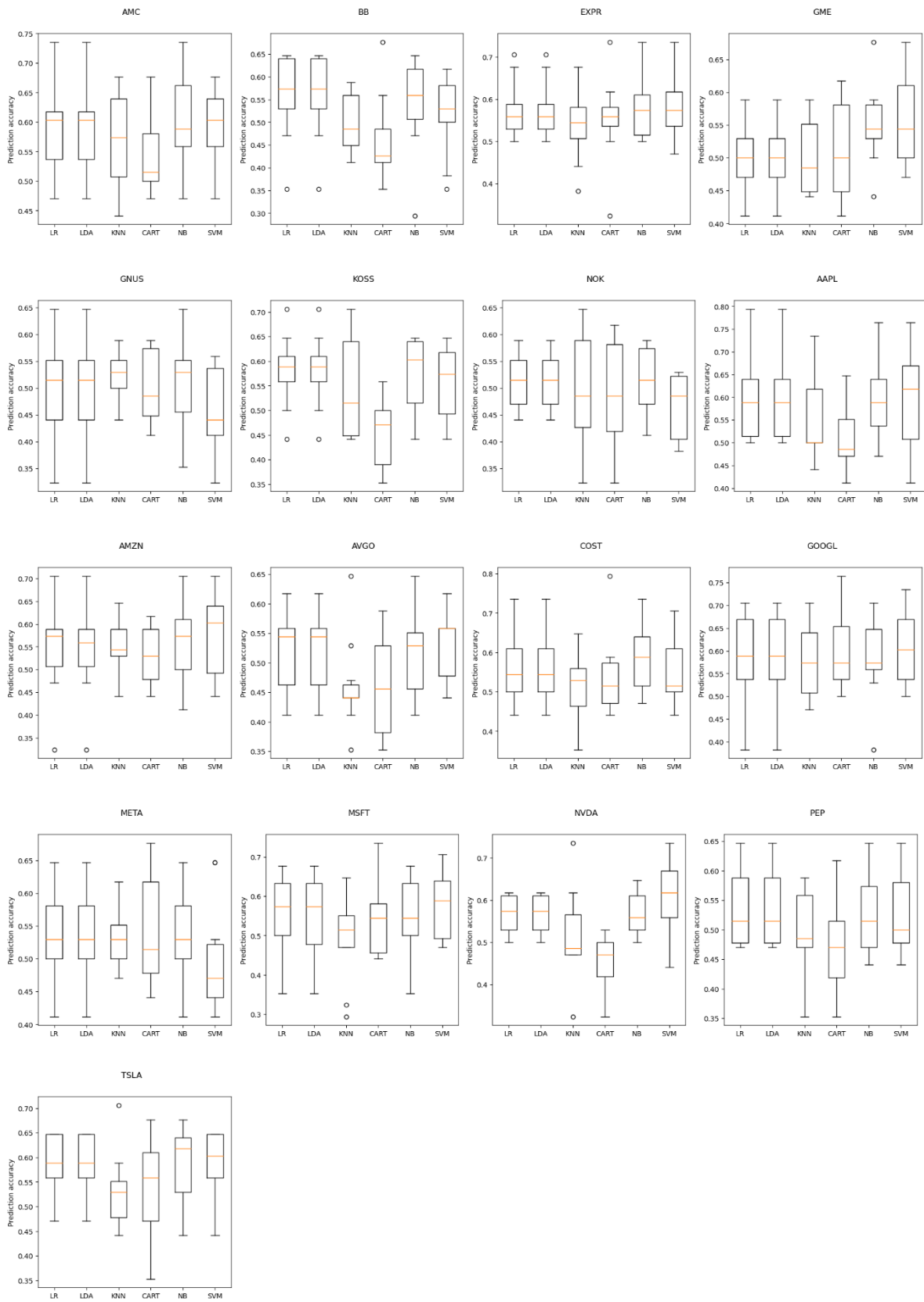
Appendix A: Lagged Correlation Plots







Appendix B: Binary Classification Plots



Appendix C: Hypothetical Portfolio Back Testing

	AMC	BB	EXPR	GME	GNUS	KOSS	NOK
Start	05/12/2019	05/12/2019	05/12/2019	05/12/2019	05/12/2019	10/12/2019	05/12/2019
End	30/09/2022	30/09/2022	30/09/2022	30/09/2022	30/09/2022	30/09/2022	30/09/2022
Duration	1030 days	1030 days	1030 days	1030 days	1030 days	1030 days	1030 days
Exposure Time [%]	69.76	84.39	68.21	72.43	80.87	72.46	91.28
Equity Final [\$]	23,949.81	18,684.15	2,105.04	8,728.33	11,133.12	12,180.90	14,239.51
Equity Peak [\$]	36,625.49	55,712.26	13,408.74	86,660.39	65,081.00	67,522.58	21,951.47
Return [%]	139.50	86.84	- 78.95	- 12.72	11.33	21.81	42.40
Buy & Hold Return [%]	- 10.89	- 12.48	- 78.20	1,470.62	90.79	350.35	23.21
Return (Ann.) [%]	36.28	24.80	- 42.44	- 4.71	3.88	7.27	13.35
Volatility (Ann.) [%]	196.44	91.49	70.97	125.60	159.15	240.78	53.00
Sharpe Ratio	0.18	0.27	-	-	0.02	0.03	0.25
Sortino Ratio	0.63	0.65	-	-	0.07	0.16	0.50
Calmar Ratio	0.48	0.36	-	-	0.05	0.08	0.37
Max. Drawdown [%]	- 75.12	- 68.25	- 84.30	- 90.27	- 82.98	- 86.35	- 35.79
Avg. Drawdown [%]	- 29.94	- 15.11	- 72.69	- 16.74	- 26.80	- 15.93	- 11.38
Max. Drawdown Duration	470 days	610 days	844 days	610 days	611 days	639 days	277 days
Avg. Drawdown Duration	112 days	90 days	513 days	61 days	121 days	69 days	62 days
# Trades	76	35	60	74	19	72	19
Win Rate [%]	43.42	40.00	31.67	39.19	42.11	55.56	42.11
Best Trade [%]	192.15	198.14	78.63	200.12	200.57	170.11	51.60
Worst Trade [%]	-33.88	-25.23	-21.35	-63.00	-32.58	-22.74	-19.03
Avg. Trade [%]	1.16	1.80	-2.56	-0.19	0.57	0.27	1.88
Max. Trade Duration	76 days	246 days	149 days	83 days	475 days	50 days	344 days
Avg. Trade Duration	8 days	24 days	11 days	9 days	43 days	9 days	49 days
Profit Factor	1.64	2.06	0.80	1.59	1.88	1.44	1.77
Expectancy [%]	3.77	5.68	-1.37	3.49	6.30	2.02	2.97
SQN	0.48	0.20	-1.26	-0.02	0.04	0.06	0.49

	AAPL	AMZN	AVGO	COST	META	NVDA	TSLA
Start	05/12/2019	05/12/2019	05/12/2019	05/12/2019	05/12/2019	05/12/2019	05/12/2019
End	30/09/2022	30/09/2022	30/09/2022	30/09/2022	30/09/2022	30/09/2022	30/09/2022
Duration	1030 days	1030 days	1030 days	1030 days	1030 days	1030 days	1030 days
Exposure Time [%]	97.61	97.47	81.86	87.20	85.23	71.73	87.06
Equity Final [\$]	25,885.61	12,013.44	18,511.10	16,018.86	6,465.97	9,796.11	58,942.81
Equity Peak [\$]	32,013.52	21,488.10	25,378.56	20,131.57	15,704.09	25,860.67	81,747.25
Return [%]	158.86	20.13	85.11	60.19	- 35.34	- 2.04	489.43
Buy & Hold Return [%]	112.11	29.85	58.85	68.85	- 31.94	133.36	1,104.33
Return (Ann.) [%]	40.09	6.72	24.39	18.18	- 14.32	- 0.73	87.53
Volatility (Ann.) [%]	49.17	38.15	39.49	27.47	36.31	43.43	125.06
Sharpe Ratio	0.82	0.18	0.62	0.66	-	-	0.70
Sortino Ratio	1.75	0.28	1.17	1.14	-	-	2.26
Calmar Ratio	1.60	0.14	0.90	0.68	-	-	1.87
Max. Drawdown [%]	- 25.04	- 49.18	- 27.06	- 26.67	- 58.83	- 62.12	- 46.81
Avg. Drawdown [%]	- 4.83	- 5.82	- 4.84	- 3.41	- 9.41	- 9.03	- 11.57
Max. Drawdown Duration	233 days	449 days	277 days	176 days	388 days	277 days	272 days
Avg. Drawdown Duration	26 days	42 days	26 days	22 days	51 days	39 days	43 days
# Trades	15	13	20	14	25	12	44
Win Rate [%]	66.67	38.46	60.00	57.14	36.00	25.00	50.00
Best Trade [%]	40.98	71.03	35.44	45.41	41.19	198.68	114.48
Worst Trade [%]	-12.30	-10.12	-15.26	-15.34	-25.11	-18.97	-16.72
Avg. Trade [%]	6.57	1.41	3.15	3.48	-1.75	-0.22	4.12
Max. Trade Duration	252 days	674 days	259 days	311 days	383 days	491 days	141 days
Avg. Trade Duration	66 days	76 days	41 days	63 days	34 days	60 days	19 days
Profit Factor	4.56	1.76	2.26	2.95	0.81	1.88	2.87
Expectancy [%]	7.46	2.94	3.86	4.36	-0.96	7.93	5.87
SQN	1.53	0.25	1.14	0.86	-0.57	-0.01	1.17