Teesside University

School of Computing, Engineering and Digital Technologies

Correlation between sentiment analysis of stock market tweets and momentum strategy in a volatile market

Isaac Dada

Supervisor: He, Mengda

A report submitted in fulfilment of the requirements of
the Teesside University for the degree of
Master of Science in *Data Science*

January 11,2023

Declaration

I, Isaac Dada, of the School of Computing, Engineering and Digital Technologies, Teesside University, confirm that this is my own work and figures, tables, equations, code snippets, in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalized accordingly.

Isaac Dada
January 11,2023

Abstract

In this study, 12500 tweets from the four big banks with the highest market capitalization were extracted from the Twitter API. The sentiment classification of these tweets was extracted using the pre-trained model RoBERTa, a predictive transformer model that outperformed earlier state-of-the-art techniques after being tuned and trained with over 124M tweets. Next, an intermediate momentum strategy was trained with stock prices for various formation dates, consecutively a specific period during the pandemic was selected to test its effectiveness. The results revealed that the strategy was able to outperform the return of the S&P 500 by 2.09% over that tested time frame. The question that now emerges is whether there is a relationship between the momentum of stock prices and sentiment analysis of large corporate tweets during those periods. Finally, the correlation between the sentiment of tweets from these four large market cap banks and the momentum rise or fall in stock prices for these winners and top performers in the intermediate momentum evaluated showed that Energy, and Petroleum companies, for example, Chevron Corporation, Occidental Petroleum were observed to have a high positive correlation with tweet sentiments as well as businesses in commodities such as oil and gas, for instance, ConocoPhillips, ExxonMobil Corporation, etc. There were also negative correlations between sentiments and stock momentum mobile communication retail companies and insurance firms, such as UnitedHealth Group Inc and Lockheed Martin Corporation among others. As the tweets extracted are unlabeled and unannotated, RoBERTa was also used for sentiment labelling, and the accuracy of this model was validated with firstly an accuracy of 0.86% as well as previous academic research using RoBERTa for sentiment classification.

Keywords: Sentiment Analysis, Momentum Strategy, RoBERTa, Efficient Market Hypothesis

Acknowledgements

My deepest gratitude goes out to He Mengda, my supervisor, for his leadership, encouragement, and support during this project. His excellent advice and suggestions have influenced and enhanced my research.

I appreciate my friends and colleagues' insightful comments and beneficial discussions. Finally, I want to express my gratitude to my family for their love and support during this time of study. This project would not have been possible if it were not for their support and confidence in me.

Contents

List of Figures

List of Tables

List of Abbreviations

CVX	Chevron Corporation
LLY	Eli Lilly And Co
PFE	Pfizer Inc.
UNH	UnitedHealth Group Inc
XOM	Exxon Mobil Corp
NOC	Northrop Grumman Corp
ABBV	AbbVie Inc
COP	ConocoPhillips
LMT	Lockheed Martin Corp
MRK	Merck & Co Inc
DVN	Devon Energy Corp
APA	APA Corp (US)
EQT	EQT Corp
MRO	Marathon Oil Corp
OXY	Occidental Petroleum Corporation
BERT	Bidirectional Encoder Representations from Transformers
AMEX	American Stock Exchange
NYSE	New York Stock Exchange
NLP	Natural language processing
DIJA	Dow Jones Industrial Average
LSTM	Long Short Term Memory

Chapter 1

Introduction

1.1 Background

Twitter as an information source

The financial market has changed over time as a result of the pandemic, and it is now more obvious than ever that participants in the financial market are becoming more and more reliant on social media platforms like Twitter (a social media platform where people can share their opinions about the market, indicators, and the movement of the market) to learn about the financial markets. Because of this, it is now more crucial than ever for technical analysts to be able to calculate the effect that news and information have on the development or decline of momentum in the stock market. The media has been inundating the public with information since the beginning of this pandemic about a wide range of subjects, including indicators and the opinions of market participants, including people and corporations, regarding various financial issues. Approximately four hundred and fifty million active monthly users are estimated to be using the Twitter in the year 2022 Dixon, S. (2022), and the number of active daily users is estimated to be approximately two hundred and thirty-seven million in the second quarter of 2022, making Twitter- one of the best places to obtain current market information. As opined by Brocardo, Traore and Woungang, (2015). However, a limitation to the use of this media resource is that it only allows an individual to post just two hundred and eighty Unicode glyphs that include icons and symbols either as text or as emojis, all of which are used to express one's feelings about a particular topic. Although, in most cases, this also accounts for noise within the data which must be cleaned for sentiment analysis to function as intended.

Intermediate Momentum Strategy

The Efficient Market Hypothesis is a subject that receives a lot of attention in the finance sector and has been the focus of many technical analysts for a long time. Fama's publication on efficient capital markets (Fama 1970), which defined market efficiency as the extent to which all pertinent information is represented in market prices, served as the initial catalyst for the phenomenon. Since there are neither undervalued nor overpriced securities accessible on the market, market efficiency indicates that all information has already been included into the pricing, making it impossible for a corporation to exceed more than the market price (Titan, 2015). The development of the momentum strategy by Jegadeesh and Titman (1993), who demonstrated that an intermediate momentum strategy produces high abnormal profits as contrasted to the market average, is the basis of a significant argument against the Fama efficient market hypothesis. They used stock data from the AMEX and NYSE between 1965 TO 1989 to rank the stocks and equities based on how profitable their returns have been over the last three to twelve months. Throughout the three to twelve months period, the objective of the test was to keep purchasing stocks from companies within the winning groups, and to continuously sell stocks from companies with low returns or losing groups, and to constantly hold them for a period of three which resulted in over twelve percent annual profit.

Furthermore, Rouwenhorst (1998) examined whether momentum existed in other markets using Jagadeesh and Titman's (1993) methodology with stocks from twelve countries in Europe with a 17 years' time frame (1978 to 1995). The research indicates that the strategy, which holds previous winning firms and sells off stocks from previous intermediate term loss incurring firms, returns around 1 percent. Furthermore, each of the 12 European countries seems to experience the momentum effect, which is not specific to one country. To buttress this, Grobys (2014) examined the returns of momentum strategies in periods of low growth. As part of the study, twenty-one foreign stock indices were chosen as a sample for the analysis, in order to implement zero-cost portfolios, grouping the indexes into various groups in terms of their previous aggregate yield.

Although the conclusion of this study differs from previous studies which had more positive returns, in that on a 10% level of significance level, Grobys' strategies displayed statistically significant negative returns during recessions. Overall, this research discovered that between 1998 and 2013, momentum-based trading methods in international equities markets were lucrative.

Roberta - Transformer Model

There is no doubt that pretraining with language models has resulted in significant performance improvements for NLP activities. However, these improvements are not without a lot of challenges, such as high computing costs, tuning hyperparameters takes time in training. In 2019, Liu et al. proposed a model based on the BERT model that Google had published the year before. Their research provided a thorough analysis of how hyperparameter adjustment and the size of the training set affected the performance of the Google's proposed model BERT. In order to improve the model's performance, Liu et al. (2019) discovered that BERT models were significantly undertrained. They suggested an improved method of training the model, which included training the model longer with larger sizes and longer sequences, removing the next sentence prediction objective, and altering the masking pattern used on the training data in ROBERTa - A Robustly Optimized BERT Pretraining Approach. It is important to note that the specific pretrained RoBERTa model - TimeLMS employed in this study is from a work by Loureiro et al. (2022), which shows a set of language models that are skilled at processing diachronic Twitter data. They used over 124 million tweets between 2019 to 2021 to train the TimeLMS model, which makes it simpler to train any additional model layer using tweets outside of its domain because the TimeLMS model was trained on a broad range of generic tweets. However, it is important to note it was quite challenging to evaluate various hyperparameters and methods because the initial 90 million tweets from the 2019 model were trained in fifteen days on eight NVIDIA V100 GPUs.

1.2 Problem statement

It is, however, still unclear how market volatility affects the effectiveness of momentum strategies for stock trading and whether sentiment analysis of tweets from major banks may help determine whether momentum develops favorably or unfavorably in the market. When it comes to putting businesses in a better position to generate higher returns, a technical analyst's capacity to quantify the effects of a particular piece of information on the movement, build-up, or decline in the momentum of a stock in the market can be crucial.

1.3 Aims and objectives

This study's goal is to evaluate the profitability of intermediate momentum strategy is during a volatile period by using the pandemic period as a case study, and to determine if there is a correlation between the sentiment analysis of tweets from high-cap companies with a focus on banks, as well as the momentum of winners within the winning decile in the momentum strategy tested. As part of the momentum strategy, stock data from the YFinance API was used, choosing a start date of 01/ 01/ 2020 and an end date of 31/ 12/ 2022 that covers for 503 S&P 500 tickers, testing 48 various formation dates, and aggregating the returns monthly. For this strategy to confirm the presence of momentum with the holding period limited to 1 month only. Lastly, to determine whether there is a relationship between tweets and comments from large market cap banks and the top performers in the performance of the momentum strategy, a RoBERTa model is utilized to create a sentiment analysis and to test for correlation, Pearson's correlation is utilized.

1.4 Summary of contributions and achievements

First, this study examined the performance of an intermediate momentum strategy within the Covid-19 pandemic period providing technical analyst with test on the link between stock momentum and market sentiments in a volatile period. Also, as it is well known that labeling financial text is expensive and requires annotators with specialized financial expertise to do it, this paper additionally contributes a non-costly approach by presenting the use of the finetuned TimeLMS for annotation of tweets sentiments. Due to the restriction of the allowed tensor size of about 514 in the model, I developed a row wise approach which is used to classify all tweets of a huge size at the same time, thus overcoming the challenge of classifying a large number of tweets at once.

1.5 Organization of the report

The rest of the paper is organized in the following manner. A comprehensive review of the literature regarding momentum strategies, transformer models, particularly the RoBERTa model for sentiment classification, as well as a number of correlation techniques and studies that indicate stock prices, volumes, and momentum are correlated with tweet sentiment is presented in the second chapter of this book. The third chapter of the study develops the methodology utilized. The results of our study, a discussion of the findings, and future work and the study's conclusion are presented in fourth and fifth chapter respectively.

Chapter 2

Literature Review

This paper is divided into several sections, including testing an intermediate momentum strategy in a volatile market with a focus on S&P 500 companies during the pandemic and analysing sentiment in tweets using the transformer model, specifically the RoBERTa model, to ascertain how momentum from winners of the intermediate momentum strategy correlates with sentiment in tweets of the time. As a result, the next section reviews prior works and studies that address the primary areas of this research.

Makrehchi et al. (2013) developed a unique method for categorizing social media text by extracting major stock market movements and gathering pertinent pre- and post-event text tweets, using significant stock market occurrences as the basis. As a result, each tweet was given a positive or negative label based on its content, and a model was trained using this collection of tweets as a foundation for anticipating the labels that subsequent tweets will be. The model showed a strong predictive power for future stock market movements based on the analysis of the net daily sentiment. Makrehchi et al. (2013) study reported that the trading approach has proven successful in outperforming the S&P 500 by roughly 20% over a period of four months.

Hutto and Gilbert (2014) proposed VADER for general sentiment analysis and evaluated its performance against 11 industry benchmarks, including SentiWordNet. The authors developed and validated a list of lexical features that are tuned to sentiment in micro blog-like contexts, and these lexical features was joined by evaluating the conditions for expressing sentiment intensity in a grammatical and syntactical manner. The rule-based model VADER surpassed individual human reviewers in terms of accuracy.

Chowdhury et al. (2014) plotted the sentiments of 15 companies over the course of four weeks to test the market's efficiency and determine whether there is any correlation between sentiment predicted by news and the original stock price. They found that on average, they were 70.1% accurate in identifying the correct sentiment. Additionally, Chowdhury et al. (2014) plotted the errors of prediction for several organizations and were successful in obtaining the Root RMSE of 30.3% and MAE of 30.04%, respectively, as well as an increased F1 score of 78.1%. Positive sentiment curves and stock price trajectories was compared, and a 67% correlation was established between them, pointing to a robust efficient market hypothesis present.

Bollen et al. (2011) examined the relationship between the values matching to mood state values produced from twitter feeds and the stock prices of DJIA with Opinion Finder, which gauges positive or negative emotions, and Mood Profile by Google. The Granger causality test was applied for the subsequent stage of the correlation analysis and for classification, a self-organizing fuzzy neural network was employed. The findings showed that by excluding the other factors as unnecessary talks and including precise and public mood characteristics.

According to Rao and Srivastava (2012) study, there is a complex relationship between tweet bullishness, volume, and agreement with market indicators like volatility, trading volume, and stock prices. With the data demonstrating a strong association (up to 0.88 for returns) between stock prices and twitter feelings. Additionally, Rao and Srivastava (2012) expressed the returns produced a high R squared with a low 1.76% Maximum Absolute Percentage Error, according to the implementation of the Expert Model Mining System (EMMS).

In Kordonis et al. (2016) developed a method that gathers tweets and evaluates the accuracy of Naive Bayes and SVM for sentiment classification. They were able to use the algorithm to evaluate the relationship in tweets and stock market price changes as well as to calculate the prediction error by contrasting the results with the actual close price the following day. According to their research, Kordonis et al. (2016) found a high average accuracy in predicting the stock's movement with a forecast error of less than 10% with reference to the expected closing price.

Pagolu et al. (2016) examined two different types of textual representations using Word2vec and Ngram and focused on the classification of sentiment in text by assuming that the opinion expressed in the text is either positive, neutral, or negative in order to analyze the public sentiments in tweets. Additionally, Pagolu et al. (2016) placed a strong emphasis on pre-processing the data through a number of filtering stages, such as tokenization, removing stop words, and regex matching for special character removal. This approach addresses one of the main issues with the previously developed analyzers, which is that they were trained on a different corpus from the one that was used for training.

Bharathi et al. (2017) approach integrated sentiment in tweets with RSS news sources and sensex stocks. The system was developed with a focus on the temporal correlation between stock market prices, sentiments expressed in tweets, and RSS feeds. For the NLP module, Bharathi et al. (2017) created a sentence-level score algorithm in conjunction with a POS Tagger and dictionary-based approach. But no comparison was made to see if these indicators could be useful as a momentum tracker; they were only taken into account based on closing price. Conclusively, their experimental study established a correlation between stock level indicators, RSS news feeds, and tweets, and demonstrated a significant 20% increase in prediction accuracy.

Das et al. (2018) developed a method for ingesting and analyzing tweet data from the Twitter API in real-time utilizing Spark streaming and Apache Flume for analysis. It is common knowledge that Spark's distributed machine learning framework may be used to process enormous amounts of data on an open-source platform called Apache Spark. Additionally, Das et al. (2018) employed open-source tools and the Lambda Architecture for the analysis of streamed data. Additionally, by using incremental active learning to adapt the methodology to a stream-based environment, their model gained the capacity to choose fresh training data from a data stream for manual labelling.

Chiong et al. (2018) by taking into account the complexity and ambiguity of natural languages used, the issue of accurately modelling stock market patterns via news releases was addressed with a support vector machine (SVM) constructed with parameters adjusted using particle swarm optimization (PSO). The experimental findings demonstrated that SVM and PSO-based prediction model's accuracy of 59.15% surpasses the 57.8% accuracy of Kraus and Feuerriegel (2017) deep learning model, which was the previous best performance.

In this study, Rodrigues et al. (2017) examined how sentiment analysis on tweets collected from

Twitter can be carried out efficiently using Apache Flume to stream live tweet data into the Hadoop Distributed File System (HDFS) sequentially using pig scripts written to extract tweets from raw twitter data into a dictionary, and then to perform sentiment analysis on those tweets. Rodrigues et al. (2017) suggested that the real-time tweet streaming be managed by the Hadoop framework, and that the Flume component interact with the Twitter streaming API to retrieve the tweets that include keywords that match the query. Therefore, Rodrigues et al. demonstrated that combining MapReduce and Hadoop was an effective technique to develop a fault tolerant processing framework. Wagh et al. (2018) have developed a general sentiment classification solution using Natural Language Toolkit where there are no labels in the target domain but some labels in a separate domain, which is referred to as the source domain, in which the data is labelled. Along with sample data that was used to train and test multiple classifiers, Wagh et al. (2018) showed a graphical demonstration that was part of the solution supplied which also utilised to highlight various results or trends. The study's findings showed that having more clean data enables the production of more precise outcomes.

Guo and Li (2019) developed a novel model of social network sentiment analysis based on Twitter sentiment scores (TSS) to predict FTSE 100 prices in real time using a baseline correlation, which resulted in a 67.22% accuracy rate on prediction and reduced computation burden, allowing for quick decision making without historical data. Rather than using tweets as a source for sentiment analysis in this study, Guo and Li (2019) used market data to develop investor sentiment, which was then compared to the linear regression model of close prices.

The work of Wu (2020) piques my interest in momentum tracking as the authors used K-means clustering with moving averages, MACD, and KDJ to select stocks based on a stock selection strategy that showed that the output had a higher excess return rate during bull markets and a rate that declined in time with the S&P500 index during bear markets.

Sagala et al. (2020) used the Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Naive Bayes among other methods in their research. The study's findings show that employing the SVM algorithm for a five-day trading window and combining features from historical data with online media sentiment led to the highest level of accuracy. Sagala et al. (2020) feature extraction method was based on the TF-IDF and Term Frequency approaches which been demonstrated that these methods can deliver exceptional results when used as the foundational building blocks for the creation of classification algorithms, particularly those based on sentiment analysis and used to forecast the movement of stock values.

With an emphasis on understanding the impact of sentiments on stock price fluctuations, Gupta and Chen (2020) explored the impact of sentiments derived from a sizable collection of tweets from StockTwits for five businesses (Apple, Amazon, General Electrics, Microsoft, and Target). Gupta and Chen (2020) used three machine learning algorithms (Nave Bayes, SVM, and logistic regression) and five feature extraction approaches to perform sentiment analysis on stockTwits data (bag of words, bigram, trigram, TF-IDF, and LSA). Based on Gupta and Chen (2020) findings, the logistic regression and TF-IDF combination was able to obtain a respectably high degree of accuracy for each of the five organizations, ranging between 75% and 85%.

With stock data from the Ghana Stock Exchange (GSE) between January 2010 and September 2019, Nti et al. (2020) tested the proposed predictive framework and made stock value predictions for time windows of one day, seven days, thirty days, sixty days, and ninety days. Nti et al. (2020) found that the accuracy ranged between (49.4-52.95%) based on Google trends, (55.5-60.05%), based on Twitter, (41.52-41.77%). With the little quantity of the data gleaned from Twitter, however, reveals that investors in undeveloped nations like Ghana rarely express their opinions on market movements on social networking services. Apergis and Pragidis (2019) study of how stock prices respond to wire news from the European Central Bank discovered a significant relationship between news sentiment tone and both volatility and the mean return on stocks. Additionally, Apergis and Pragidis (2019) showed that during times of crisis, the correlation between news emotions and stock market volatility became stronger. The analysis creates a new index that captures the mood of the public's reaction to these statements during the 2008 financial crisis.

For the study of product reviews, Dai et al. (2021) suggested the RoBERTa-IAN transformer model-based sentiment analysis model. A pre-training model called RoBERTa was utilized to convert the context and aspect words of product reviews into low dimensional vectors. Then, in order to extract semantic information from the low dimensional vector and create a hidden representation based on it, the low dimensional vector is fed into the Bi-GRU model. In order for Dai et al. (2021) to assess and categorize the sentiment polarity of product reviews, they employed an Interactive Attention Networks to collect the attention matrices of context and aspect words. These matrices were then transferred to a sentiment classification layer for additional classification. According to experimental findings using a real commodity dataset from China, RoBERTa-IAN was found to have a 90% accuracy rate in predicting commodity prices.

In a study that was published by Ghasiya and Okamura (2021), they used a database of headlines and articles to examine more than 100,000 COVID-19 news articles and headlines using top2vec for topic modelling and RoBERTa for sentiment classification and analysis. The topic modelling results showed that some of the most popular and widely covered issues are sports, economy, and education. Furthermore, the validation accuracy of the Ghasiya and Okamura (2021) sentiment classification model was 90%, and the analysis showed that the UK was the worst-affected nation and had the highest percentage of negative sentiment of 73.23%, while South Korea had the highest percentage of positive news of 54.47%. Tan et al. (2022) proposed a hybrid deep learning approach in this study that suppresses the limits of the sequence model while combining the benefits of the Transformer model and sequence model. To capture long distance context each word is mapped into a small, word embedding space using RoBERTa and LSTM. They found that the model performs better than the existing approaches, with F1-scores of 0.93 and 0.90 on the IMDb, and Sentiment 140. Furthermore, the SentiSmoke-Twitter and Reddit datasets, as well as an extensive annotation schema for identifying the sentiment of tobacco products, were released as a result of a study conducted by Yanamandra et al. (2021) that investigated the sentiment and product identification of texts relating to tobacco on two social media platforms. Tan et al. (2022) used models like BERT, RoBERTa, and DistilBERT in their bench marking text classification studies.

Researchers from all across the world have been fascinated by the resilience and key drivers of momentum in the years since Jegadeesh and Titman (1993) first discovered it. According to Rouwenhorst (1999) research, return momentum can also be seen among companies listed on emerging markets' stock indexes, which supports Rouwenhorst (1998) empirical study that showed momentum tracking strategies are lucrative. His study showed that there is medium-term return continuity in global equities markets. After adjusting for risk, a portfolio of historically diverse winners beat a historically diversified losers in all twelve study countries between 1980 and 1995, this effect lasted, on average, for around a year.

By purchasing equities depending on their return throughout the ranking period, Chan et al. (2000) used a weighted relative strength strategy (WRSS) to compare the performance of 23 countries over a three-year period. The authors found that momentum techniques are successful when used in the global equities markets, which is consistent with Rouwenhorst (1999) findings.

The most frequently mentioned arguments for the existence of momentum profit, according to Tse (2015), are regarded to be initial underreactions of stock prices and delayed overreactions later on, even if the causes of momentum profit have not yet been fully established. Tse (2015) investigate the profitability of momentum strategies utilising the sector exchange-traded funds (ETFs) traded on the NYSE and the iShares exchange-traded funds (ETFs) as individual investors can benefit from the stock market's movement in both directions by engaging in short sales and margin purchases, and the ETFs facilitate speedy entry into the market. Tse (2015) discovered that employing relative strength performance techniques, no appreciable momentum profits were seen for any combination of ranking and holding periods.

The profitability and return structure of momentum strategies are examined by Abourachid et al. (2017) across ten European nations using data from December 2004 to December 2015. Abourachid et al. (2017) discovered that momentum returns were smaller than those mentioned in other studies. The results showed that market conditions from the height of the financial crisis to its aftermath, from 2007 to 2012, can be blamed for the low momentum returns. Additionally, Abourachid et al. (2017) discovered that the momentum returns produced by small and large stocks over distinct time periods were stronger for small stocks and worse for large ones. Finally, the author discovered that 10 of the 16 techniques that skip one month between the formation and holding periods produce statistically significant returns throughout the whole sample period.

The association between intermediate-term momentum and credit ratings is examined by the author using a sample of 4447 Standard & Poor (S&P) credit rated stocks from December 1984 to December 2011 as a proxy for credit risk, Haga (2015). The author demonstrates that intermediate term momentum is substantial for a wide range of enterprises independent of the firms' credit ratings, supporting Novy-Marx (2012) claim that intermediate-term momentum is more resilient than short-term momentum. Highlights the fact that only high and medium credit rating enterprises have different intermediate and short term momentum.

In Gray and Vogel (2016) which evaluated the efficient market theory' recommendation and the perspectives of technical and quantitative analysts on momentum strategy to determine the relevance of momentum in stock selection strategy. According to Gray and Vogel (2016), intermediate momentum strategy outperforms short and long momentum strategy with a performance difference of about 4% annually. The most important finding was the emergence of a clear trend: keeping fewer stocks and re-balancing more regularly results in better CAGRs.

When market indices rose between 1930 and 1932 it produced large losses, as shown by Jegadeesh and Titman (1993) back-testing of momentum strategies ; nonetheless, comparatively few studies have been conducted to determine whether momentum techniques are beneficial during economic downturns. While Daniel and Moskowitz (2013) found that momentum portfolios exhibit a strong up- and down-beta differential during bear markets especially mostly the portfolios of losers. Chordia and Shivakumar (2002) found that momentum payoffs appear to be negative during recessions but they are not statistically different from zero during those periods. As a result, if market conditions improve, these losers will see significant gains indicating a fall in momentum.

According to Brenøe (2018) research, machine learning can produce abnormal returns by taking use of the well-established momentum of time series patterns that have been demonstrated to exist in the past. Lagged and Hedged momentum strategies, as well as portfolio configurations like WML trading techniques on hedge funds, were used as part of the testing to evaluate the accuracy of various ML models. The study showed that the portfolio with the most winners also contain a lot of losers, and vice versa. It was also found that both portfolios contain a sizable percentage of stocks that are regarded as neutral in character Brenøe (2018).

Chapter 3

Methodology

3.1 Momentum Strategy

I used the Yahoo Finance API (the API is free and allows scraping of stock data for a specific period of time) to extract stock data from all 503 S&P 500 Ksu (2022) companies during the period of time from January 1, 2018, to December 31, 2022, in order to build the momentum strategy. The momentum strategy approach was initially developed using the extracted data's adjusted pricing for the dates mentioned above. The momentum strategy is predicated on the sum of returns over the previous twelve months, a methodology that also adheres to the conclusions of Jegadeesh and Titman (1993). Our holding period was one month before we gave any refactoring. In order to strengthen their study, the method was additionally examined with a one-month interval between the ranking and the holding period. The acronym K/S/J, which represents the momentum strategies variables, stands for the number of ranking months the company stocks, the number of holding months, and the number of months between ranking and holding. In this study, our K value corresponds to 12 months, a J value to one month, and a S value to one month. Finally, the return for each company's stocks can be represented by Jegadeesh and Titman (1993) formula:

$$R^{j}_{i,T} = \pi^{j}_{t=1}(1 + R_{i,T-(t-1)}) - 1. \tag{3.1}$$

Following a careful analysis of all the returns for the 48 formation dates, it was determined that 2021 should have the highest percentage of high returns and that September represent the highest return within the high. Thus, 2021 was selected as the focus year for the remainder of the study, which included the time period for tweets extracted from the Twitter API.

3.2 Sentiment Analysis with Roberta

3.2.1 Data Preprocessing

I utilized the Twitter Developer portal API to obtain 12500 sample tweets from four users of focus in particular the four largest banks based upon market cap, namely JPMorgan, Goldman Sachs, Citi and BoA Business between 2019 and 2022. To retrieve all previous tweets, we performed a simple API query that ignored retweets and pulled all previous tweets. However, there was a limit of 2500 tweets per user due to limitations of the twitter developer API regarding data size. Moreover, Twitter only allows you to extract tweets counting backwards from the date the function was executed because it has depreciated the functions: "start from" and " end date", so the limit has been placed in order to prevent the extraction of unneeded tweets that are outside the range of interest. As a next step, in order to get just tweets that are related to the financial market and not a generalized corpus of tweets, I used the list of 151 keywords from Apergis and Pragidis (2019). Thus, a smaller sample of 2724 tweets were selected for further analysis that matched the stated keywords. The data cleaning process involved removing all stop words from the tweets that were contained within the tweets using a Natural Language ToolKit library, which is a combination of all stop words along with essential neutral words that don't have a significant impact or addition to the sentiment analysis within the tweets. The second thing that I did was to convert all tweets text to lowercase, remove all hyperlinks to other web pages, remove punctuation, and remove all next line characters white space that appeared in the tweets using regex pattern replacement. As part of the data pre-processing process, as the demoji package was recently depreciated, I had to create a function that removed all emotion icons, all pictographs, all symbols, all non-English characters and flags, by evaluating their Unicode representation, and finally, as stipulated in Apergis and Pragidis (2019) In order to improve the accuracy of sentiment analysis, all combinations of found "@" and "usernames" were replaced with a general "@user".

3.2.2 Architecture and Implementation

A modification of the robustly optimised BERT pretraining method, which is based on the transformer model (BERT), is employed as part of the sentiment classification approach. Compared to the initial 90M base tweets utilized by Liu et al. (2019), the Roberta - TimeLMS I used was trained on 124M tweets (Loureiro et al., 2022). In comparison to the current machine learning models employed in the field of natural language processing, the RoBERTa model demonstrated superior performance. A longer training period, more training data, bigger batches, and longer sequences were employed to train RoBERTa, which is a more robust variant of BERT. The main distinction between this model and its predecessor - BERT, is that the new model was trained exclusively with dynamic masking with the masked words or tokens changing within various training epochs whereas the old model was trained with next sentence prediction. We had to import a configuration file and weights for the model from hugging face, which already had and stored the pretrained model and tokenizer. However, the tweets sample were divided into four separate users, and each separate sentiment classification was carried out for each user separately to account for the computational cost of running the model. The classification procedure took about 2 hours and 40 minutes to complete in comparison to a much longer time with combined tweets, however this was primarily because of the amount of the sample data used in the analysis. I used a SoftMax function on the outputs based on the logits to convert the results into probabilities, on which I was able to classify the sentiments into 3 groups: neutral, negative, and positive based on their probabilities split between 0 and 1. The highest probability was assigned as the most likely sentiment for each tweet that was processed using some data manipulation techniques Eq. (3.2), Fig. 3.1. For validation, I trained the RoBERTa model on 11k tweets across 4 epochs using the same pre-trained model weights, and when utilizing validation data for validation, the classification produced a log loss decrease for each epoch and an improvement in accuracy of 0.87. However, because it took more than 6 hours, this was also computationally expensive using a NVIDIA Standard NC6 6 cores, 56GB RAM virtual machine learning studio.

$$=INDEX(\$A\$1:\$C\$1,SUMPRODUCT\quad(MAX((\$A\$2:\$C\$251=E2)*(COLUMN\quad(\$A\$2:\$C\$251)))-COLUMN\quad(\$A\$1)+1) \tag{3.2}$$

A	B	C	D	E	F	G
Negative	Neutral	Positive	Tweet Content	Most Likely	Result	
0.025943	0.811521	0.162536	agile payments infrastr	0.81152076	**Neutral**	
0.001168	0.037968	0.960864	city lights exhilarating I	0.96086377	**Positive**	
0.001921	0.247319	0.75076	enchanting city inspire:	0.75075984	**Positive**	
0.008927	0.752064	0.239009	machine learning cybe	0.75206435	**Neutral**	
0.008096	0.425706	0.566199	oneonone mentoring I	0.5661986	**Positive**	
0.009952	0.615705	0.374343	innovation meets insig	0.61570495	**Neutral**	
0.017931	0.553063	0.429006	jennifer piepszak belie	0.5530633	**Neutral**	

Figure 3.1: Data Manipulation Result

3.3 Correlation between Sentiments and Stock Momentum

3.3.1 Stock Momentum Indicator - MACD

The moving average convergence divergence oscillator (MACD), a technical momentum indicator, is used in this methodology to distinguish between the short-term (12 days) and long-term (26 days) exponential moving averages (long term) as shown in the eq. (3.5). One can better understand the extent to which momentum varies, either positively or negatively, over a specific period of time, by comparing the aforementioned exponential moving averages with one another. A higher 12-day exponential moving average suggests a positive momentum, whereas the contrary shows a negative momentum. In order to calculate MACD, Signal Lines, and MACD Histograms, the following formulas were utilized.

$$12 - dayShortEMA = ticker.Close.ewm(span = 12, adjust = False).mean() \tag{3.3}$$

$$26 - dayLongEMA = ticker.Close.ewm(span = 26, adjust = False).mean() \tag{3.4}$$

$$MACD = 12 - dayShortEMA - 26 - dayLongEMA \tag{3.5}$$

$$signal = MACD.ewm(span = 9, adjust = False).mean() \tag{3.6}$$

$$Histogram(Momentum\ rate) = MACD - signal \tag{3.7}$$

3.3.2 Pearson's Correlation Heatmap

As part of the process of testing the correlation of the extracted sentiments, we renamed them from neutral to 0, negative to -1, and positive to 1 before using a Pearson correlation heat map to examine the correlation fig. 3.3. Following that, the results of the sentiments analysis calculated for all four aforementioned users were correlated to the MACD as shown in the figure 3.2 below.

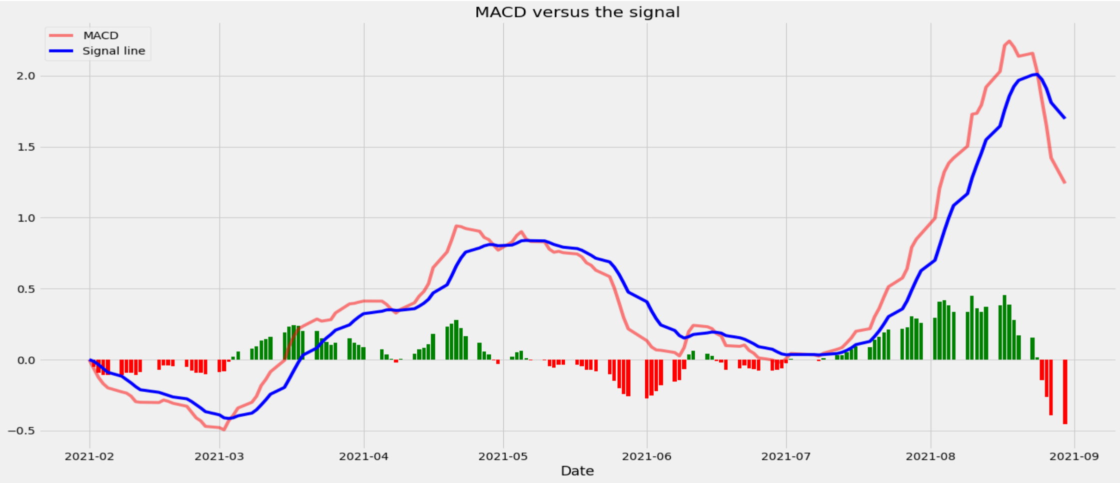

Figure 3.2: MACD Indicator

For the purpose of testing the above results, two hypotheses were defined and filtering out of the correlation with a p-value less than 0.05, the correlation coefficient findings were tested for statistical significance fig. 3.4. The null hypothesis should be rejected if the p-value is less than 0.05. If the p-value is greater than 0.05, then the null hypothesis is not rejected.

NULL HYPOTHESIS H0: There is no statistically significant relationship between the sentiment of tweets and the momentum of the stocks (in this case is represented by the MACD which is the different between the short term and long-term EMA).

ALTERNATIVE HYPOTHESIS H1: There is a statistical significance between the correlation of sentiment of tweets and the momentum of the winners in the strategy utilized.

However, when the correlation was tested based on correlation between the sentiments of tweets and the difference between the MACD and signal line (which indicates the degree of acceleration of positive or negative momentum across stock prices), the results in fig 3.5 showed that the COP, the DVN, the APA, and the EQT were negatively correlated. Moreover, it should be noted that the correlation fig 3.5 is only valid for items that have a p-value < 0.05, thus the null hypothesis is rejected. For correlation items with a p-value > 0.05, as seen in fig 3.4, no statistically significant correlation was found, thus correlation results were masked for all stocks.

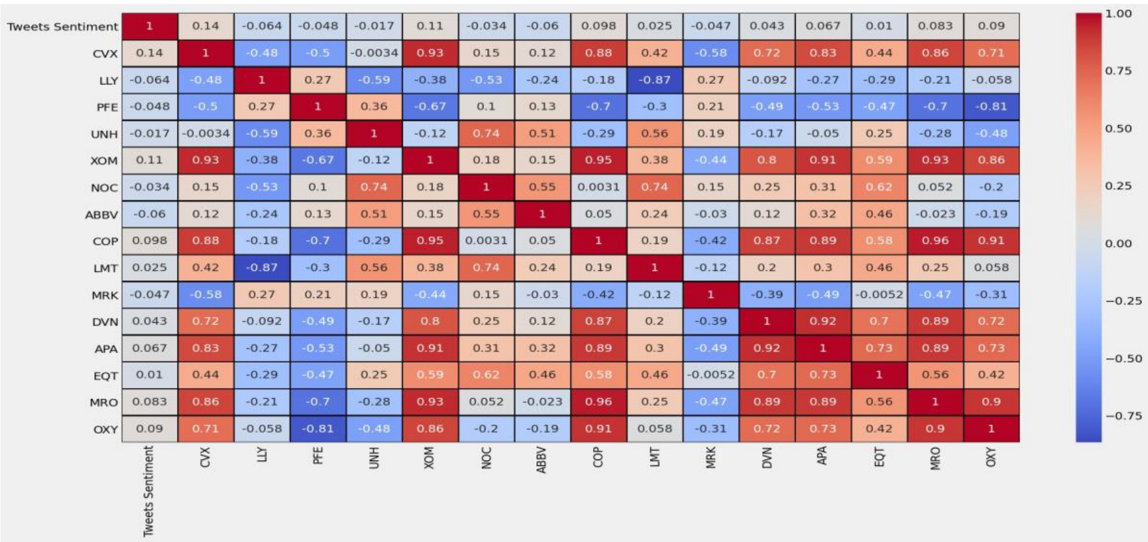

Figure 3.3: Correlation Heatmap

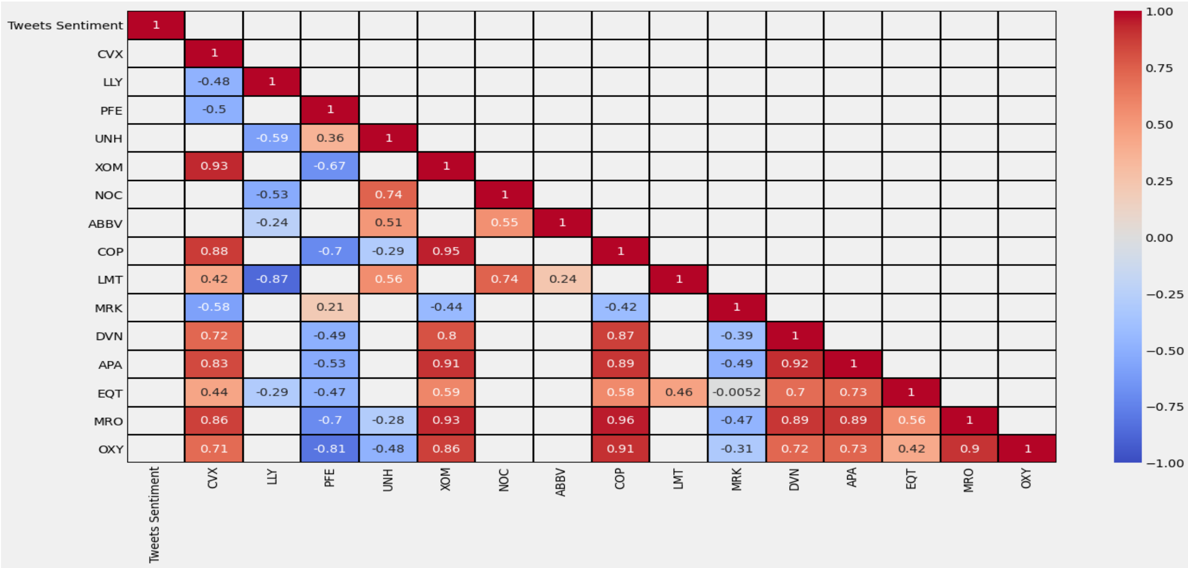

Figure 3.4: Statistical Significance Testing with MACD

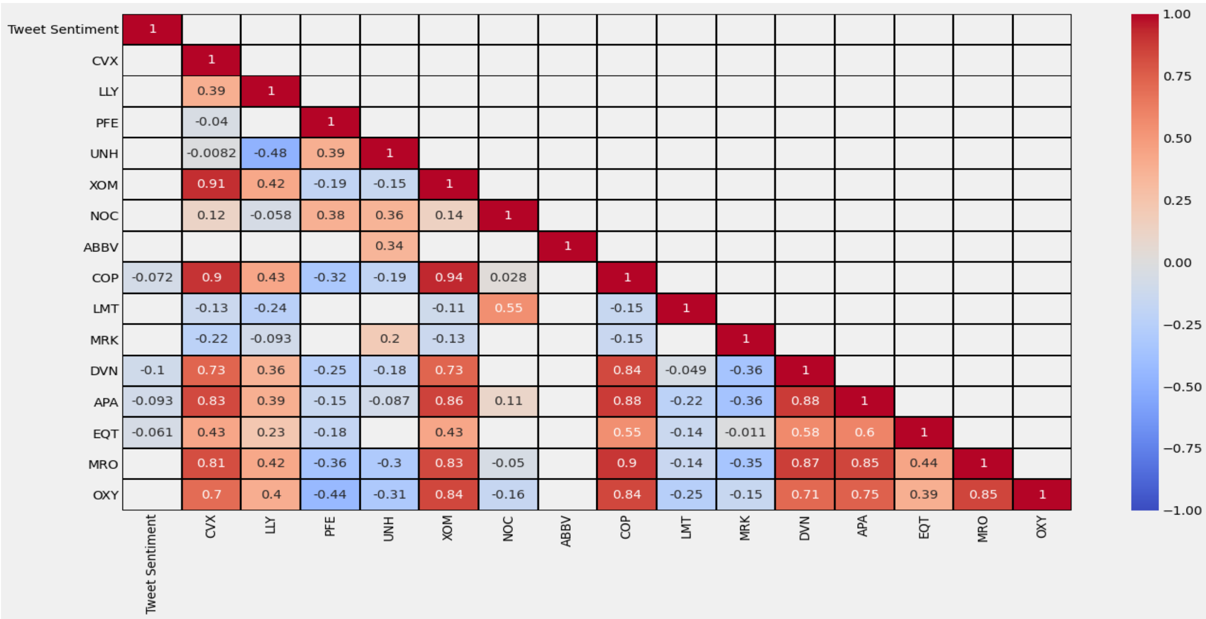

Figure 3.5: Statistical Significance Testing with Momentum Strength

In analyzing all the momentum winners that were considered in the correlation, including companies with the highest market cap, the top five winners by their highest return, and the sentiment classification of tweets among those companies, there was a 100% positive correlation between the top five winners by their highest return and tweet sentiment classification fig 3.3 This indicates that the sentiment of tweets is quite the same as the direction of momentum of the stock's prices of the companies with the highest return. On the other hand, a positive correlation of 40% was found for the top 10 companies by market capitalization.

Chapter 4

Results

Previously, it has been stated that the purpose of this study was to investigate the effect of market volatility on the effectiveness of intermediate momentum strategies for stock trading, as well as the effect of sentiment analysis of tweets from major banks in determining whether momentum develops favorably or unfavorably in the market. Here are some of the results that were found.

Results from the intermediate momentum strategy:

- The initial test produced 51 winners in the winning decile. The top 5 winners by return on profit were the energy, oil, and hydrocarbon corporations: DVN-Devon Energy Corporation, OXY-Occidental Petroleum Corporation, MRO-Marathon Oil Corp, EQT-EQT Corporation, and APA-APA Corporation. The top 10 largest winners by market capitalization (Moore, 2022): UNH (UnitedHealth Group), XOM (ExxonMobil Corporation), CVX (Chevron Corporation), LLY (Eli Lilly and Co.), PFE (Pfizer Inc.), ABBV (AbbVie Inc.), MRK (Merck Co. Inc.), COP (ConocoPhillips), LMT (Lockheed Martin Corps.), and NOC (Northrop Grumman Corp.) are also all a mix of pharmaceutical, aerospace defense and energy corporations.

- For 48 different formation date tested, there was a profit return for 27 formation dates with an average profit return of 5% across these dates. There was also recorded 21 days non profitable negative returns averaging at 6% across these days as shown in the table 4.2. In 2019 and 2020, there is a significant and positive average return ranging from 0.01% to 1.15% as well as an average negative return of -0.06% in some months. Also, the number of positive monthly returns was higher in 2021, with over six formation dates and eleven total samples, and the average return ranged from 0.01% to 0.05%. It is evident that the intermediate momentum strategy performs exceptionally well for the entire period of 2019-2022, with a positive return of 56.25% for all formation date samples examined

- The momentum strategy, however, consistently beat the S&P 500 market returns from September 2021 to August 2022, on an average by 60% of the tested times.

Table 4.1: All formation dates tested

SN	Profitable Dates	Non-Profitable Dates
1	01/02/2019	01/01/2019
2	01/03/2019	01/04/2019
3	01/05/2019	01/06/2019
4	01/07/2019	01/09/2019
5	01/08/2019	01/10/2019
6	01/01/2020	01/11/2019
7	01/02/2020	01/12/2019
8	01/03/2020	01/04/2020
9	01/05/2020	01/06/2020
10	01/07/2020	01/08/2020
11	01/09/2020	01/11/2020
12	01/10/2020	01/12/2020
13	01/01/2021	01/02/2021
14	01/04/2021	01/03/2021
15	01/05/2021	01/06/2021
16	01/08/2021	01/07/2021
17	01/09/2021	01/11/2021
18	01/10/2021	01/12/2021
18	01/10/2021	01/12/2021
19	01/02/2022	01/01/2022
20	01/03/2022	01/07/2022
21	01/04/2022	01/11/2022
22	01/05/2022	
23	01/06/2022	
24	01/08/2022	
25	01/09/2022	
26	01/10/2022	
27	01/12/2022	

Table 4.2: Comparison of intermediate momentum strategy returns

Dates	Momentum returns	S&P 500 returns	Outperformed
30/09/2021	0.055124	-0.01191	Yes
31/10/2021	-0.013055	0.069144	No
30/11/2021	-0.07166	-0.008334	No
31/12/2021	-0.044191	0.043613	No
31/01/2022	0.008194	-0.052585	Yes
28/02/2022	0.055483	-0.03136	Yes
31/03/2022	0.020926	0.035773	No
30/04/2022	0.057099	-0.087957	Yes
31/05/2022	0.020976	0.000053	Yes
30/06/2022	-0.064901	-0.08392	Yes
31/07/2022	0.062449	0.091116	No

Sentiment classification with RoBERTa:

Using the transformer model - RoBERTa model, I was able to classify sentiment in tweets from four of the largest banks based on their market capitalization. With 86% validation accuracy, our model classified the tweet's text better than other traditional classifiers. The total number of tweets that were pulled from the Twitter API was 12500 with 2500 tweets assigned per unique users queried, with the earlier date range largely influenced by the volume of daily posting carried out by these companies. When compared to other companies, those with a higher volume of tweets within the specified period range has its oldest tweets close to the present date. See the Table 4.3.

Table 4.3: Extracted Tweets

User	Most Recent Post	Oldest Post
BofA-Business	26/10/2022	23/08/2017
Citi	26/10/2022	29/08/2019
GoldmanSachs	26/10/2022	21/05/2020
WellsFargo	27/10/2022	01/09/2022
jpmorgan	26/10/2022	20/02/2018

The result of the sentiment analysis before the data manipulation to select the maximum sentiment per tweet classification was carried out showed that there was a lesser degree of negative sentiments within the tweets classified as seen in the figures below. In line with the previous methodology described in this paper, in my implementation of the RoBERTa model for sentiment classification, I found the following results: out of the 456 filtered tweets from JPMorgan that had keywords with market focus, 69.89% were neutral classified tweets, while 30.11% were positive classified tweets over the specified time frame. Among the 698 filtered tweets from GoldmanSachs, there were 0.80% negative classified tweets, 74.80% neutral classified tweets and 24.40% positive classified tweets, thus indicating that GoldmanSachs has the highest neutral class of tweets, and it is the only company among the four companies that has negative classified tweets. In the 664 filtered BoA-Business tweets, 59.80% of the tweets were classified as neutral, while 40.20% of the tweets were classified as positive. There were 47.85% neutral classified tweets in the 550 filtered Citi tweets, and 52.15% positive classified tweets, making Citi the user with the highest percentage of tweets with positive sentiments. It is important to note that WellsFargo tweets were not processed for sentiment classification as the time range in which the extracts were taken was between 01/09/2022 and 27/10/2022, which is way out of the time frame of 2021 considered.

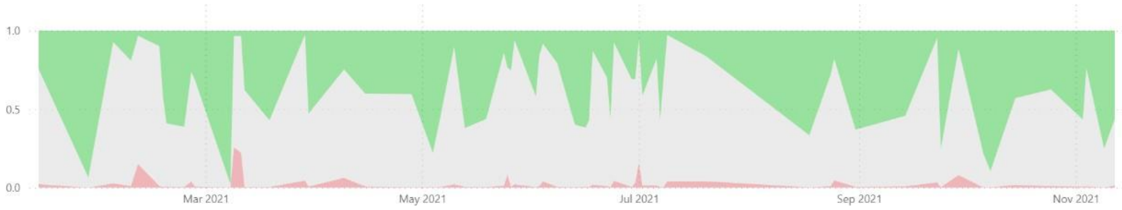

Figure 4.1: JPMorgan Sentiment Results

Figure 4.2: Citi Sentiment Results

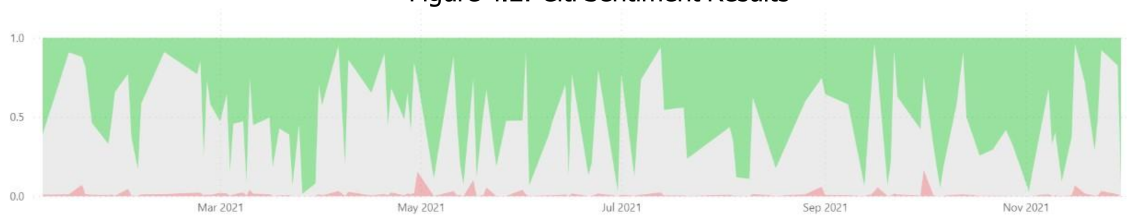

Figure 4.3: GoldmanSachs Sentiment Results

Figure 4.4: BoA Business Sentiment Results

Correlation results: Two methods of correlation were used: correlation between MACD line and sentiment classification and correlation between MACD and signal line difference and sentiment classification, which yielded varying results. In order to accept or reject these hypotheses, both were tested using p-values < 0.05. See Figure 3.3 and Figure 3.4

- In the first approach, positive correlations were observed for CVX, XOM, COP, LMT, DVN, OXY, MRO, EQT, APA, and negative correlations for LLY, PFE, UNH, NOC, ABBV, MRK. However, the constraint of p-value < 0.05 was not enforced. For hypothesis testing, no correlation coefficient was found when the p-value < 0.05.

- In a second approach, CVX, PFE, UNH, NOC, LMT, MRK were all positively correlated with MACD and signal line difference, while XOM, COP, DVN, OXY, MRO, EQT, APA LLY, ABBV, and MRK were negatively correlated. While the hypothesis testing revealed a statistically significant negative correlation in COP, DVN, APA, and EQT and the difference between MACD and signal line when p value < 0.05 as seen in Figure 4.4. Therefore, the null hypothesis is rejected.

Chapter 5

Discussion and Analysis

5.1 Significance of the findings

It has been shown by various aforementioned momentum strategy research ranging from Jegadeesh and Titman (1993) to Brenøe (2018) that momentum strategies are profitable in the financial markets, and the result of the intermediate strategy in this study also shows that momentum was present in the market as well as being profitable. Although a lot more months within the formation date of the strategy returned profits as shown in table 4.1, however months with losses impacted the average yearly returns significantly. Therefore, the profitability of the strategy was not clearly reflected within the pandemic period.

With a validation score of 86%, the sentiment classification results were consistent with other research using Roberta. The results ranged between 80-90% Dai et al. (2021), Ghasiya and Okamura (2021) depending on the number of epochs completed with log loss decreasing with each epoch, but it is computationally expensive to continue training. Also, there was a higher distribution of positive results of 35.97% and neutral results of 63.71% as compared to negative results which was less concentrated at 0.32%. This was due to the limitation of not having enough text to classify, since Twitter only allows users to post 140 characters for each tweet. As a result, most of these large banks provide a more in-depth account of their thoughts through a hyperlink that leads to their website, where restrictions are more limited.

To evaluate the statistical significance of our correlation, the p-value, which is calculated together with the Pearson coefficient, can be read as follows: The correlation is regarded as statistically significant when the p-value is less than 0.05, and the estimated Pearson coefficient can be relied upon. The correlation, however, is not regarded as statistically significant when the p-value is greater than 0.05 and may have happened randomly. Hence the study's correlation findings indicate that for some of the intermediate momentum winners, there is a significant negative correlation between stock momentum and sentiment analysis of 26.67%. However, it differs slightly from earlier studies like the method of Chowdhury et al. (2014), which showed a 67% positive correlation between sentiment and stock prices, and the study of Apergis and Pragidis (2019), which examined how stock prices reacted to wire news from the European Central Bank, which found a significant relationship between news sentiment tone and both volatility and return. This approach differs from mine in that it focuses on finding the effect of tweets from specific financial institutions as opposed to general tweets or well-known market drivers like the European Central Bank which is likely to have a more general effect on market decisions.

5.2 Limitations

It would be challenging to cover all tweets from all of the commercial banks and companies in the study, so the study was restricted to four significant banks with sizable market capitalization. The limitations of our study are discussed in this section.

- The maximum tokenizer length on the tokenizer configuration has been set by the RoBERTa model to 514, which means that a tensor input greater than this value won't be able to be processed by the tokenizer. In order to get around the restriction, reset max length to a smaller value than the allocated value.
- Data collection also faced limitations, such as the requirement that developers have access to twitter API, and the difficulty in specifying a specific date range that is appropriate to scrape from twitter API as this function is now depreciated.
- Due to the limited computational power available, we only used a small sample of data to classify tweets because processing a huge dataset required a substantial amount of time. In this instance, just tweets from four large banks were considered because including more prolonged the computation.

It's important to keep in mind that the particular domain of tweets and the time period being taken into account may have an impact on the strength and direction of the correlation between tweet sentiment classification and stock price momentum. The findings of our study indicated that sentiment analysis can be a useful method for forecasting stock momentum and offers vital insight to technical analyst. However, as there are many variables at play in the stock market, it is important to use caution when relying on these results.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This study examined the correlation between sentiment analysis of tweets from large market cap banks specifically JPMorgan Chase, Goldman Sachs, Citi, and BoA Business and stock momentum from winners in an intermediate momentum strategy. The intermediate momentum strategy involved buying stocks that have recently outperformed the market and selling those that have underperformed after holding for a period before refactoring. On the other hand, sentiment analysis in- volved analysing the emotional tone of tweets text focused specifically on the market in order to gauge public opinion.

The momentum strategy was found to have a 56.25% profitability rate over the course of 48 formation dates. Additionally, it was observed that with an 86% validation rate, the RoBERTa outperforms accurately in predicting and annotating sentiments of tweets. Also, there was a significant negative correlation between stock momentum and sentiment analysis for the stocks of ConocoPhillips, Devon Energy Corp, APA Corp (US), and EQT Corp which are major players in the energy and hydrocarbon exploration market. The purpose of this study was only to examine the correlation between stock momentum and sentiment analysis for these specific stocks and test for the significance of this relationship. However, the study did not attempt to establish a causal relationship between them since a wide variety of factors and conditions play an important role in determining which direction momentum rises or decreases either positively or negatively on the financial market.

6.2 Future work

In future works, I want to try to get beyond some of the limitations of this study on the basis of a critical assessment of the findings in section 4. This includes increasing the number of tweets and text sources, with a larger data sample size would undoubtedly having an impact on the precision of our RoBERTa model. This would include adding tweets and text sources for more companies, especially banks that have a large impact on the market, such the US Federal Reserves. Additionally, it would be beneficial if I tried to combine RoBERTa's model with other neural networks to boost its overall accuracy. This study has demonstrated that intermediate momentum methods and sentiment analysis in the stock market can perform well together as a technical indicator for the market. However, in order to comprehend the connection between stock momentum and sentiment analysis in various financial contexts and for a number of different equities, more research is required.

References

Abourachid, H., Kubo, A. and Orbach, S. (2017). Momentum strategies in european equity markets: Perspectives on the recent financial and european debt crises. *Finance Research Letters*, 23:147–151.

Apergis, N. and Pragidis, I. (2019). Stock price reactions to wire news from the european central bank: Evidence from changes in the sentiment tone and international market indexes. *International Advances in Economic Research*, 25(1):91–112.

Bharathi, S., Geetha, A. and Sathiynarayanan, R. (2017). Sentiment analysis of twitter and rss news feeds and its impact on stock market prediction. *International Journal of Intelligent Engineering and Systems*, 10(6):68–77.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Brenøe, M. (2018) Applying artificial intelligence on momentum. 2018.

Brocardo, M.L., Traore, I. and Woungang, I. (2015) 'Authorship verification of e-mail and tweet messages applied for continuous authentication', *Journal of Computer and System Sciences*, 81(8), pp. 1429-1440.

Chan, K., Hameed, A., and Tong, W. (2000). Profitability of momentum stragegies in the international equity markets. *Journal of financial and quantitative analysis*, 35(2):153–172.

Chiong, R., Fan, Z., Hu, Z., Adam, M., Lutz, B., and Neumann, D. (2018). A sentiment analysis-based machine learning approach for financial market prediction via news disclosures. In *Proceedings of the genetic and evolutionary computation conference companion*, pages 278–279.

Chordia, T., and Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *The journal of Finance*, 57(2):985–1019.

Chowdhury, S., Routh, S., and Chakrabarti, S. (2014). News analytics and sentiment analysis to predict stock price trends. *International Journal of Computer Science and Information Technologies*, 5(3):3595–3604.

Dai, J., Pan, F., Shou, Z., and Zhang, H. (2021). Roberta-ian for aspect-level sentiment analysis of product reviews. In *Journal of Physics: Conference Series*, volume 1827, page 012079. IOP Publishing, 2021.

Daniel, K., and Moskowitz, T. (2013). Momentum crashes. *Swiss Finance Institute Research Paper*, (13-61):14–6.

Das, S., Behera, R., and Rath, S. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia computer science*, 132:956–964.

Dixon, S. (2022) Twitter: Number of users worldwide 2024, Statista. Available at: https://www.statista.com/statistics/303681/twitter-users-worldwide/ (Accessed: January 11, 2023).

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), pp.383-417.

Ghasiya, P., and Okamura, K. (2021). Investigating covid-19 news across four nations: a topic modeling and sentiment analysis approach. *Ieee Access*, 9:36645–36656.

Gray, W., and Vogel, J. (2016). *Quantitative Momentum: A practitioner's guide to building a momentum-based stock selection system*. John Wiley Sons, pages 12-235.

Guo, X., and Li, J. (2019). A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 472–477. IEEE.

Gupta, R., and Chen, M. (2020). Sentiment analysis for stock price prediction. In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 213–218. IEEE.

Haga, J. (2015). Intermediate-term momentum and credit rating. *Finance Research Letters*, 15:59–67.

Hutto, C., and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Jegadeesh, N., and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.

Kordonis, J., Symeonidis, S., and Arampatzis, A. (2016). Stock price forecasting via sentiment analysis on twitter. In *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, pages 1–6.

Kraus, M., and Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *arXiv e-prints*, page arXiv: 1710.03954.

Ksu. (2020). List of S&P 500 companies, 2022. URL https://en.wikipedia.org/wiki/List_of_S%26P_ 500_companies.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019) Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Loureiro, D., Barbieri, F., Neves, L., Anke, L., and Collados, J. (2022). Timelms: Diachronic language models from twitter. *arXiv preprint arXiv:2202.03829*.

Makrehchi, M., Shah, S., and Liao, W. (2013). Stock prediction using event-based sentiment analysis. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 1, pages 337–342. IEEE.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, 103(3): 429–453

Nti, I., Adekoya, A., and Weyori, B. (2020). Predicting stock market price movement using sentiment analysis: Evidence from ghana. *Appl.Comput.Syst.*, 25(1):33– 42.

Pagolu, V., Reddy, K., Panda, G., and Majhi B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pages 1345–1350. IEEE.

Rao, T., and Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis.

Rodrigues, A., Rao, A., and Chiplunkar, N. (2017). Sentiment analysis of real time twitter data using big data approach. In *- 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, pages 1–6.

Rouwenhorst, K. (1998). International momentum strategies. *The journal of finance*, 53(1):267–284.

Rouwenhorst, K. (1999). Local return factors and turnover in emerging stock markets. *The journal of finance*, 54(4):1439–1464.

Sagala, T., Saputri, M., Mahendra, R., and Budi, I. (2020). Stock price movement prediction using technical analysis and sentiment analysis. In *Proceedings of the 2020 2nd Asia pacific information technology conference*, pages 123–127.

Tan, K., Lee, C., Anbananthen, K., and Lim, K. (2022). Roberta- lstm: A hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, 10:21517–21525.

Tse, Y. (2015). Momentum strategies with stock index exchange-traded funds. *The North American Journal of Economics and Finance*, 33:134–148.

Wagh, B., Shinde, J., and Kale, P. (2018). A twitter sentiment analysis using nltk and machine learning techniques. *International Journal of Emerging Research in Management and Technology*, 6(12):37–44.

Wu, S. (2020). Application of cluster analysis in stock selection in united states stock market. In *Pro- ceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning*, pages 310–313.

Yanamandra, V., Pant, K., and Mamidi, R. (2021). Towards sentiment analysis of tobacco products' usage in social media. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1545–1552.

Appendix

The list of keywords used in filtering the tweets are as follows:

algo,anti-city,anti-dilution,anti-speculation,anti-speculative,arbitrage,arbitrageur,asset management, asset-stripping, bagger, bear, bear market, bearish, blue chip, bond, bourse, broker, bull ,bull market, bullish, buy something in, buyback, callable, capital investment, carpetbagger, cash, cash out, cashout, city, copper-bottomed, crash, bounce,dealing,delist,dilution,disinvest,disinvestment,distributable,diversified,dividend,Dow Jones, electricals, equity, faller, Fibonacci ,flotation ,footsie ,fund management ,fund manager, future ,gainer ,gilt ,go public idiom ,growth-oriented ,grubstake ,haircut ,head fake, hedging, holding ,ICO ,industrial ,initial coin offering ,insider dealing ,intangible asset, investment ,issue ,list ,listed company ,mature ,maturity ,negotiable ,non-directional, non-discretionary, non-distributable, non-rated ,non-speculative ,noncallable ,nondiversified, OFEX, option ,outgain ,over-investment ,par ,payback period ,plan ,portfolio ,post-crash, principal ,public ,pyramid scheme ,pyramid selling ,quant ,redeemable ,rentier ,repurchase, rig, rig the market idiom ,rogue trader ,security ,seed money ,share ,shareholder ,small cap, sound ,speculation ,speculative ,speculatively ,spread betting ,stag ,stake ,stakeholder ,stock, stock exchange ,stock market ,stockbroker ,stockbroker belt ,stockbroking ,stockholder ,strip, subscribe ,tender ,the Big Board ,the FTSE 100 ,the grey market ,the Nikkei index ,the S&P 500, S&P 500 ,the Square Mile ,trade ,trader ,trading ,Treasury bond ,unissued ,unit trust, unlisted ,venture capital ,venture capitalist ,Wall Street, administered prices , employment , inflation , aggregate demand , energy price , inflation , inflation rate , borrowing growth , energy prices , inflationary pressure , broad money , equity prices , inflationary pressures.

A note file is attached that gives a description of all datasets used in the artifact.