Geneva Business School®
Real business. Future leaders.

Research Topic

Measuring the Impact of Individual Twitter Accounts on the S&P 500 and Investment Strategies

Master Thesis
Geneva Business School

Master in International Finance

Submitted by:

Michelle Fleming

Geneva, Switzerland
Approved on the application of:

Dr. Samer Ajour

Dr. Oliver Elliot

And

Dr. Roy Maurad

Date: May 22, 2020

Declaration of Authorship

"I hereby declare:

- That I have written this work on my own without other people's help (copy-editing, translation, etc.) and without the use of any aids other than those indicated;
- That I have mentioned all the sources used and quoted them correctly in accordance with academic quotation rules;
- That the topic or parts of it are not already the object of any work or examination of another course unless this has been explicitly agreed on with the faculty member in advance;
- That my work may be scanned in and electronically checked for plagiarism;
- That I understand that my work can be published online or deposited to the university repository. I understand that to limit access to my work due to the commercial sensitivity of the content or to protect my intellectual property or that of the company I worked with, I need to file a Bar on Access according to thesis guidelines."

Date: May 22, 2020

Name: Michelle Fleming

Signature: Michelle Fleming

Acknowledgements

I would like to thank all of the faculty and staff at Geneva Business School for giving me hands on tools and real world experiences during my masters program. I would especially like to express my gratitude to my advisor Samer Ajour for being a dedicated teacher and mentor through this process.

Table of Contents

Abstract

With the stock market seeing some of its most volatile days in history, it is more important than ever for investors to be able to reduce and mitigate risk for their investments (Froot, Scharfstein & Stein, 1993). There is the constant push for higher returns and more accurate predictions from investors as well as from higher up executives on analysts and portfolio managers in order to reach target earnings. Due to these increased pressures, investors may feel the need to take on more risk in their portfolios than they normally would in order to achieve their benchmarks. The need for better foresight is always a sought after attribute for those looking to invest in portfolios or individual companies.

From previous research, it has been alluded that the social media platform, Twitter, can be used as a short term stock market indicator with relative accuracy (Rao & Srivastava, 2012). It has been recorded that the social media platform had an accuracy rate of nearly 90 percent from a 2011 study looking at the daily up and down changes for the Dow Jones closing values (Mao & Zheng, 2011). While in the past, studies have been focused on the social media platform as a whole, our study looked at the link between the accounts of influential individuals and movements in the market. The results of this paper have come to the conclusions that personal Twitter accounts can have the ability to be used as a stock market indicator for the S&P 500 as well as a way to craft better performing portfolios.

From the results, there is a link between observed variables relying heavily on the concept of sentiment analysis and retweet count per trading day. With an accuracy rating of 91.2 percent in the predictions based on the Ordinary Least Squared (OLS) Model and the ARIMA time series analysis model, there are enormous opportunities for the financial industry to implement these findings in order to enable smarter and less risky investments. With future developments in enabling Twitter extraction, we believe that individual twitter accounts can be used in order to craft better performing portfolios in the market, mitigating risk and outperforming conventional methods.

Chapter 1: Introduction

Investing is inherently risky (Litner, 1975). As there is no way yet to be able to tell the future, investors will continue to take risks both making and losing money with each purchase or sale they execute (Weber, Weber & Nosic, 2013). While there are ways to minimize risk while investing, such as diversifying your portfolio by investing in a variety of sectors with low correlation and using hedging strategies to offset losses (Bodie & Marcus, 2004), even with complex financial risk management tools (Van Deventer, Imal & Mesler, 2013) there is no way to reduce risk to zero percent (Lakonishok, Shapiro, 1986). There will always be systematic risk that cannot be prevented (Alaghi, 2012) within the market and is impossible to reduce with the implementation of diversification (Jordan, Ross, Westerfield, 2004). Risky stocks tend to have higher volatility and therefore higher payouts as a reward for taking on the additional liability (Hunt, Moyer, Shevlin, 2007).

Risk exposure can be critical to the downfall of a portfolio and even large financial institutions (Flannery, 2017). Companies taking on excessive risk and a lack of regulation regarding risk profiles were a few of the major reasons for the financial crisis in 2008 (Baker, 2008). Although it is difficult to determine risk reward profiles (Mainelli, 2004), it is imperative to implement strategies to manage and control risk while looking at investments (Bower, 1972) for both individuals as well as institutional investors (Cardona, 2013).

If it were possible to be able to predict the movements of the stock market before they were to happen, risk would ultimately be reduced (Greenwald & Stiglitz, 1993). Individuals lacking experience or knowledge about the stock market would be more inclined to participate if risks were lower and payoffs were higher and more likely (Holt, Laury, 2002). Traditionally, methods such as surveys were used in order to collect and interpret information (Williams, 2007). Not only can surveys be extremely costly, but they can also have issues such as responder truthfulness, individual and social biases, as well as group think (Mao & Bollen, 2011) which will be discussed later in the paper. With new technology and widespread innovation around the world, it may now be possible to forecast future stock market movements with relative accuracy through the use of platforms such as social media (Nyugen & Shirai, 2015). This paper will be looking at the feasibility of using individual Twitter accounts to anticipate movements in the market. A study conducted in 2011, came to the conclusion that Twitter has up to "87.6 percent [accuracy] in predicting the daily up and down changes in the closing values of the DJIA" (Mao & Zheng, 2011). Building on the analysis and studies of Mao and Zheng's foundational work could potentially change the way individuals and firms go about making investments in the stock market.

Twitter is a social media platform where users are able to convey thoughts and opinions in concise microblogs called tweets (Java, 2007). Founded in 2006 in the tech hub of the United States of America in San Francisco, California, it is currently estimated to have 134 million daily active users which are constantly providing the platform with new material (Lin, 2019). When Twitter was first established, tweets were limited to only 140 characters, but in 2017 the limit was changed so that users could post up to 280 characters, enabling users to state more complex and insightful thoughts on the platform. Although Twitter users pale in comparison to that of Facebook or Instagram, which both boast over 1 billion active users, it is still heavily influential in the social media realm and for the possible observation of financial market movements (Bakshy, Hofman, Mason & Watts, 2011) .

Twitter is a social media platform dedicated to keeping people informed and reflecting changes almost immediately (Twitter.com, 2020). Users tweet what they feel or think at that time, not later on in the day. It is a platform of current opinions and ideas. On many platforms, individuals may only post once a week or even less, but active Twitter users may post multiple tweets per day. During the impeachment trials in January 2020, president Donald Trump is recorded to have tweeted 142 times in one day, setting the record for the highest tweet count for a president (Cummings, 2020).

In addition, other social media platforms are mostly used to showcase a person's "best" version of themselves, such as instagram or Facebook, while Twitter has more of a reputation for being the more cynical or a "real" social media platform where users reflect the type of self they would present in a real social situation (Mueller, 2018). People tweet what they are experiencing or feeling at that moment and others are able to relate to them at the same time. Tweets from even a few days ago are regarded as irrelevant in the Twitter world because so much new content has already been published. The short-term nature of the platform makes it a good indicator of short term stock movements (Mao,Wei,Wang & Liu, 2012). Thus, by observing the public's opinion on certain topics, investors are able to determine general price movements from positive or negative tweets.

Investing strategies using social media data will be most effective in the short term because of the previously mentioned life-span of the content, but market and investment strategies differ from country to country (Blomstrom, Kokko & Zejan, 2000). There are over 60 major stock exchanges that exist around the globe with over 69 trillion dollars traded yearly (World Bank, 2019). The majority of money traded can be separated into three major regions: North America, Europe and Asia, making up 93 percent of global transactions (Martin, 2019). The largest stock exchange by far, is the

New York Stock Exchange, which has an estimated average trading size of 28.5 trillion USD (NYSE.com). In comparison, the second-largest stock exchange, also based out of the United States, the NASDAQ, trades on averages about 7.5 trillion dollars annually (Desjardins, 2016). The stock market is the driving force behind the world's economy and it is instrumental in maintaining growth around the globe, especially for developed economies which rely heavily on services and innovation in technology rather than manufacturing or production (Kellerman, 1985).

While the market is the driving force behind many of the major economies (Mussa, 2000) around the world, it also experiences many ups and downs throughout any given day. It has undergone some extreme movements throughout history, most recently at the time of writing with the largest Dow Jones Industrial Index (DJIA) point drop in history during the Coronavirus pandemic in March 2020 of nearly 3,000 points (Millhiser, 2020). This happened four days after the previous record downturn of 1,200 points, also due to Coronavirus fears (Menton, 2020). With threats of a recession approaching in the approaching months and the on-going trade war between the United States of America and China (Hughes, 2005), the stock market has been experiencing unprecedented volatility over the past year (Chong & Li, 2019). Additionally, not even a year before, in August of 2019, the Dow Jones Industrial Average experienced its fourth largest all-time drop of just above 800 points, over a 3 percent decline, after bond prices reflected a strong indication of an approaching recession (Li, 2019).

Because of the erratic and volatile nature of the stock market (Garber, 1996), it is necessary to use various types of instruments in order to ensure the best strategy is being utilized (Stenger, 2003). There are different financial instruments that can be used depending on the type of risk an investor (Bender, 2013) is willing to take on as the market encompasses different types of transactions (Fabozzi, 2008). These transactions range from over the counter (OTC) transactions (Dodd, 2012), requiring a broker, such as the purchasing and selling of publicly traded stocks, to items such as financial securities including exchange traded funds (ETFs), derivatives, corporate bonds and commodities (Duffie, Gârleanu & Pedersen, 2005). Stock exchanges bring together buyers and sellers from all over the world (Petram, 2014). In the past, brokers were needed in order to carry out most transactions over the phone, but today nearly all trading is automated with little need for physical human interaction (Hendershott & Moulton, 2011). The market is generally believed to be fairly priced and transparent for both buyers and sellers making it efficient for both parties (Malkiel, 1989).

The Efficient Market Theory, first mentioned by Eugene Fama in 1970, states that markets are impossible to predict and stock prices reflect all available information presented at a period of time for companies (Fama, 1970). This theory is governed by random events and is believed to be extremely difficult to outperform the market consistently. There are three levels of the hypothesis ranging from weak to strong, where different amounts of public and private information are thought to be reflected in the prices. The theory operates under the hypothesis that investors act rationally when presented with new information. While there have been thousands of studies conducted after the initial paper titled Efficient Capital Markets: A Review of Theory and Empirical Work (1970) was published, with general support of the theory, over the past few decades, many economists and researchers supporting the notion of behavioral finance and value investing have come to find the theory highly unlikely (Brown, 2020).

Behavioral finance is a new approach to conventional finance (Griffith, Najand & Shen, 2020) where the idea is that financial information cannot be reflected accurately in the market because some agents or investors do not act rationally (Barbaris & Thaler, 2003). With heavy influences from psychology and biases from customers beliefs or fears, groups of investors acting on these feelings can have long-term effects on the markets without accurately demonstrating the true value of companies or stocks (Kashif, Shaik & Rehman, 2020). The Efficient Market Theory believes that prices in the market are correct and there is no room for arbitrage, or the "investment strategy that guarantees a positive payoff in some contingency with no possibility of a negative payoff, and with no net investment" (Dybvig & Ross, 2008). With this belief derived from those findings, there would be no way to make money from price discrepancy quotes between listed prices and true values of stocks (Williams & Dobelman, 2020).

Additional research conducted on the link between psychology and investing have led to the notion that people have a tendency to overreact when faced with unexpected or startling news (Wei, 2016). "People tend to overweight recent information and underweight prior (or base rate) data" (De Bondt, 1985) leading to large movements between prices without a justified basis. Aligning with the concept of behavioral finance, dramatic behavior in response to current news will have more drastic effects on the outcome of short-term movements (Ahmad, 2020) in financial markets compared to actual company performance or anticipated returns (Gross, 2013).

The paper entitled, Investor Psychology and Asset Pricing, written by David Hirshleifer in 2001, states that psychology based asset pricing theories are still in their infancy (Hirshleifer, 2001). While there is supporting evidence for both sides of the efficient market theory and the psychological approach, he also comments on the inability for either experienced or novice investors to be immune from mispriced assets. The general notion follows the idea that revenues flow from inexperienced traders to more skilled investors because of the gap of knowledge between the two. Within the finance industry, analysts and investors are believed to be able to predict stock movements with more accuracy than an average person outside the field (Hirshleifer, 2001). “When substantial mispricing is limited to a few factors and residuals, less rational investors do not necessarily lose on average to wiser ones. Investors who underestimate risk take larger, long positions in risky assets, and thereby achieve higher expected returns” (DeLong, 1991). In reality, the so called “experts” are not accurate in their predictions that much more often than inexperienced traders, but often they are able to manipulate the markets in the way they want because of the large amounts of money they are able to invest under the corporate umbrella (Benabou & Laroque, 1992).

There are additional theories related to psychological investing and behavioral finance which also go against the idea of the Efficient Market theory. The idea of value investing, the investment strategy which is most famously backed by self-made billionaire, Warren Buffett, is used as a mechanism to benefit from price discrepancies for stocks due to “anomalies” (Chan & Lakonishok, 2004).

The strategy relies on one of the most popular analysis tools, the fundamental analysis of companies, where “the use of current and past financial statements in conjunction with industry and economic data in order to determine a firm’s intrinsic value and identify mispriced securities” (Kothari, 2001). Technical analysis, which looks at past performance such as price movements and trading volume to indicate future performance, is not as useful for this strategy because technical analysis focuses on looking at price graphs and using trend reading to predict short term future performance (Lee, 2020). If the Efficient Market Theory were true, arbitrage events would not be possible to be achieved by investors (Brown, 2020) and long term strategic investments would be nearly impossible because assets would already be efficiently priced. Value investing is a longstanding strategy where investors purchase highly undervalued stocks and wait for the markets to reflect the accurate prices for stocks before selling (Piotroski, 2000).

Investing in the market could be thought to be easier than ever with new developments in trading technology such as free online financial technology (fintech) platforms including E-Trade or Robinhood, and access to information by the public (Baker, Nofsinger & Puttonen, 2020). Often because there is so much information available it is challenging to weed out accurate, unbiased reporting (Woo, Mai, McAleer & Wong, 2020). With only the pop-up of a push notification from a news station, investors can be easily swayed by topics chosen to be covered by the media. With never ending coverage of stories on political decisions, climate change, technological developments and scandals, the markets are heavily influenced and altered by media stories around the world (McGregor & Molyneux, 2020). Because media is selective in the stories they choose to print or air, it is difficult to determine the causal effect of the media on financial markets (Engelberg & Parsons, 2011). It is more probable that there is bias based on news stations relating to political affiliation and news anchors personal beliefs whether intended or not, such as the stock market reaction to Covid.19 public health interpreted as a political cause (Baker, Bloom, Davis, Kost, Sammon & Viratyosin, 2020).

Over the past 15 years, social media has created a new landscape for financial markets, as what individuals choose to share or like is not controlled by the government in the majority of countries (Liu, Xia & Xiao, 2020). Social media is defined by Webster's dictionary as a "form of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content" (Dijkmans, Kerkhof & Beukeboom, 2020) Social media has continued to gain users since its inception and has evolved into a more mainstream form of communication (Wang, Yu & Wei, 2012). It is estimated that nearly 4.5 billion people have access to the internet worldwide, with approximately 3.6 billion social media users reported in 2019 (Clement, 2019). With so many active users around the globe expressing their opinions, social media is able to capture the "wisdom of the crowd" and has been linked to being a good indication of firm equity value based on web-blogs and customer ratings (Lou & Zhang, 2013). With this more relevant and up to date information coming from social media accounts, it may be useful for analysts looking at the profitability of companies to make more informed decisions on their selections (Audrino, Sigrist & Ballinari, 2020).

With the high probability of the Efficient Markets Theory being incorrect (Fender, 2020), it may be possible to use alternative methods in order to predict financial outcomes in the market with unprecedented accuracy (Madura, 2020). This paper will focus on the effects of individual social media accounts, specifically Twitter, and the influence of their tweets on the stock market. The individuals observed have a large following on the app with a high number of likes and retweets everytime the app is used. Individuals with status, such as politicians, celebrities and CEOs usually have a greater influence in comparison to the average person (Levin, 2020), which is why the research will also focus on their personal accounts as opposed to company Twitter accounts. Studies have been conducted over the past two decades observing the correlation between Twitter and the stock market in general, but there is a gap in the literature looking into the individual effects of personal accounts on the market.

Currently, it is extremely difficult to predict short term stock market movements with accuracy besides having good luck and being very skilled at reading stock price trends and catching the exact right time to buy or sell (Zhou, Gao, Liu & Xiao, 2020). Very few people have these skills and it requires years of practice and often losing a lot of money in mistakes in order to become an expert. With additional tools, those less experienced may be more willing to engage in the stock market and more capital will be injected into financial markets. With further study and development, there is the hope that Twitter could be used for more accurate stock analysis for better crafted portfolios. In the literature review it will be seen how Twitter can be used as a stock market indicator because it is a platform used to express feelings through the assessment of sentiment analysis (Mittal & Goel, 2012), is applicable and reliable in the short term (Rao & Srivastava, 2012), and can influence large movements because of herd mentality as individuals are largely swayed by public figures and opinions (Dhang & Lin, 2016) and can be considered financial innovation (Mihet & NYU, 2020).

The paper is organized in four sections. The first section will provide a better understanding of the current literature of the sentiment analysis and its effect on the stock market mechanism. The methodology used in this thesis is explanatory and will use the inductive method in order to test a new hypothesis, and thus the second section will be addressed using a quantitative methodology to check on the impact of twitter sentiments on the stock market prices. The third section will reveal the findings, and the fourth section will conclude and raise awareness of the need for further research concerning behavioral financial markets.

Chapter 2: Summary of Literature Review

Social media platforms such as Facebook, Instagram, Sanpchat and Twitter have been linked to a variety of factors influencing individuals' lives including people's state of mental health and can even lead to more drastic effects such as anxiety and depression (Woods, 2016). With such great effects on an individual's mental health, it is questionable whether or not social media can have an impact on additional aspects of people's lives such as financial well being (Chen, 2013). Recently, social media has been a hot topic for its influence over the potential outcome of elections and politics in general, having been most widely looked at for the Russian involvement in the United States 2016 presidential election (Allcott & Gentzkow, 2017). But, this poses the question of whether it is possible for social media to have an influence on the financial aspects of society including stock market movements or if social media's influence is restricted to its influence on human behavior. The current literature encompasses how social media platforms specifically focusing on Twitter could be used as a short term stock market indicator. With a spotlight on Twitter as a whole, the literature indicates with strong conviction that Twitter can be used as a stock market indicator (Porshnev,Redkin & Shevchenko, 2013) but leaves gaps relating to the effects of individual accounts.

Studies conducted around the world over the last few decades have established a link between tweets on Twitter and stock market performance with a generally positive correlation (Mao & Zheng, 2011). This indicates that Twitter could potentially be used as a way to predict stock market movements and possibly even a way to craft better portfolios that may outperform the market as opposed to utilizing conventional methods of analysis. Although the general Market Efficiency Theory, stating that markets are impossible to predict and stock prices reflect all available information presented at a period of time (Fama, 1970), has been widely accepted by investors over the past 50 years, social media influence and widespread availability of information and data may be a discreditation to this phenomenon. The current literature covers topics including sentimental analysis, the importance of applying social media trends during certain time periods (Rao, 2014) as well as the psychological aspect of group think influencing individual behavior based on group mentality (Janis, 1972). Isolated cases demonstrating individual influences through the use of Twitter in the financial markets in the past are included as well.

In one study conducted by two PhD students in 2010, it was estimated that Twitter had the ability to predict the market with an accuracy of up to “87.6 percent in forecasting the daily up and down changes in the closing values of the DJIA and a reduction of the Mean Average Percentage Error (MAPE) by more than 6 percent” (Mao & Zheng, 2010). With nearly 90 percent efficiency, users could potentially feel more confident in their investing decisions and the result of the outcomes.

Along with the accuracy of predictions, using Twitter for investment decisions has also been linked to producing better results compared to solely using historical analysis (Smith, Fisher & Youngjian, 2012). In general, using historical analysis is not the most effective way to trade stocks, as past performance is not an indicator of future success (Ro, 2016). In another journal published in 2015, there was a study conducted where traders using social media for their investments averaged “2.07 percent better returns compared to solely using the historical model” (Nguyen & Shirai, 2015). Establishing better return for investors and a high probability of accuracy is an influential outcome to persuade investors to look more closely at Twitter trends and overall sentiments when making investment decisions (Zhang, Fuehres & Gloor, 2011).

Companies have already started to create algorithms based on Twitter and sentiment analysis, for specific stock market prediction (Nan, Perumal & Zaiane, 2020). Chicago-based leader in the predictive and market intelligence sphere, Social Market Analytics Inc, founded in 2012 (socialmarketanalystics.com) has made it their mission to aid day traders and other short term investors with their sentiment reading algorithm.

With a greater adoption of this type of technology across the board, the author offers it as a way to ensure more steady returns, lower volatility in portfolios and a larger profit for both individuals and firms. There is always a need to mitigate risk when investing in the stock market (Kahn & Baum, 2020), especially dealing with pension or mutual funds where hard working people rely on financial investors to secure funds for their retirement. With better analysis and higher returns, there will be less risk taking by financial managers to provide unrealistically high returns and be more beneficial for the financial industry as a whole (Menkhoff, Schmidt, Brozynski 2006).

Sentiment Analysis:

Sentiment analysis is the main tool used by researchers in order to look at and observe short term stock movements (Shapiro, Sudhof & Wilson, 2020). Sentiment analysis "is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language" (Liu, 2012). It is a useful method to determine public opinion as stock shifts in the short term are generally related to speculation (Alexander, 1961). It is extremely difficult for individual traders to accurately predict these movements in the market using conventional methods such as technical analysis because "most big institutions use computers to automate their trading, and retail traders cannot compete with algorithms daily because it is too difficult" (Deo, 2019). Therefore, utilizing sentiment analysis is one possible effective method of approaching investments that would result in less speculation as well as more reliable outcomes (Kouloumpis, Wilson, Moore, 2011).

With the constant stream of new content from Twitter, it is especially beneficial for analysts to observe current events and news trends on the platform as news has a drastic influence on stock prices (Chen, Lazer, 2013). New developments in company performance such as over or under estimations of quarterly expectations or using public relations tools to manage executive actions such as scandals are almost immediately reflected in stock prices (Mitchell, Mulherin, 1994). While these discoveries have negative implications for stock quotes, they may not necessarily have long term negative implications for the future of the company.

For example, when Steve Jobs died in 2011, Apple stock dropped by more than 5 percent (Kollewe, 2011). This result does not reflect any real impact of the company valuation of Apple but rather reflects external factors and investors' opinions about the influence of Steve Jobs on the company. Today, Apple stock has continued to be a ground-breaking market leader with continued large gains since Job's death and was the first company to achieve a market valuation of over one trillion dollars (Shell, 2018). Although Steve Jobs was instrumental in the development of Apple as a company, his death did not mark the peak of innovation or financial returns as the company continues to reach new milestones.

In addition to general sentiment analysis, researchers have looked into determining the distinctions between posts that are positive and negative and the various effects the tweets have on the outcome in the market (Oliveira, Cortez & Areal, 2017). It has been shown that positive and negative tweets do have different relationships between stock market movements. From the literature, it can be gathered that researchers believe there is a distinct difference between varying feelings such as hope and fear corresponding to positive and negative emotions (Lopes,1987). There appears to be less discrepancies between feelings that are often grouped together on the emotional spectrum such as happiness and hope. While tweets that have an optimistic outlook might have a somewhat beneficial impact on a company, negative tweets with harsher sentiments will have a much more drastic impact on stock market movements. (He, Gou, Shen & Akula, 2016). As the previous results demonstrate, people tend to be swayed more by negative news instead of positive news.

With news having large effects on the mentality of investors, it can be daunting for risk averse investors when the markets and specific stocks start to act more volatile. Risk averse individuals are more worried about the possible loss of their investments and are more hesitant about injecting their money into stocks with high uncertainty (McDonald, Siegel, 1986). Because of this nervousness, they will often react quickly when news is released about the possible underperformance or downturn of a stock or dividend payments (Eades, Hess & Kim, 1985).

Psychologists Baumeister, Bratslavsky, Finkenauer and Vohs state in their paper published in 2001 that “bad emotions, bad parents, and bad feedback have more impact than good ones, and bad information is processed more thoroughly than good” (Baumeister, 2001). Because our minds have stronger associations with negative outcomes, tweets reflecting negative sentiments such as fear or anger tend to have more drastic reactions in comparison to positive hopeful or happy tweets.

From the previous literature, it can be gathered that Twitter is generally able to accurately reflect user sentiments at that time in history and this makes it an effective tool for gauging public opinion on various topics ranging from politics to company reputations (Makazhanov, Rafiei & Waqar, 2014). With such a high track record of accuracy in stock market prediction, it can be determined that sentiment analysis is one of the most powerful mechanisms used in order to look at short term stock market movements (Kooloumpis, Wilson, Moore, 2011). Through tracking positive and negative tweets from the general public through the use of hashtags, words used to group associated topics together with the use of the # symbol, it could be a useful tool for analysts in order to predict stock movements (Lim & Butine, 2014) even more so than solely using general analysis such as looking at financial statements or conventional news outlets.

While sentiment analysis may be difficult to determine (Rojas-Barahona, 2016), it is one of the most effective when observing Tweets in order to look at overarching market feelings from the public. Users are generally very vocal while using their social media platforms and it is a way to observe true unprejudiced feelings about a specific topic or company from an outside perspective. Having access to a wide array of opinions both with and without financial knowledge allows for observations with less bias as opposed to looking at opinions from solely within the business industry from those trained in the field (Gonzalez-Bailon, Wang, Rivero, Borge-Holthoefer & Morean, 2012).

Short Term Analysis:

There are two ways to invest in the market. An individual can choose to have a short term perspective or a long term perspective depending on their needs or requirements. Often for long term investing, individuals choose to look at companies with steady dividends, a good company reputation and they look for stocks which they think are undervalued (Bodie, Kane & Marcus, 2003). Individuals may also invest in bonds, where returns are significantly lower but risk is much less or even zero, for example when investing in stable countries government bonds such as the United States or Germany.

Short term investing can be a bit more erratic, sometimes with what seems to be little to no rationale behind stock market movements. While there are more reasonable ways to look at long term investments, short term investing is more difficult and very easy to lose a large percentage of a portfolio very quickly (Froot, Scharfstein & Stein, 1992).

Investing in the stock market for the short term is highly affected by what is happening in the news and stock prices can be radically altered based on breaking news releases or by scandals. There is a “window of influence” which indicates a time interval that news will have an effect on a stock (Gidófalvi, 2001). Movements due to speculation have very little to do with how the company is actually doing or the profitability for the future but rather perceptions surrounding the company as well as political or social climates (Tumarkin & Whitelaw, 2001). Because news changes everyday, so do stock prices and there are price discrepancies between prices reflected in the market and actual fair value for stocks.

Day trading is an especially difficult aspect of the stock market as there are often many fluctuations up and down for stocks in the market (Poterba & Summers, 1984). Technical analysis, or the trading method "on the basis of largely visual inspection of past prices, without regard to any underlying economic or 'fundamental' analysis" (Taylor, 1992) is extremely challenging for investors because they must be able to pick up on visual cues and ever changing trends in the market nearly instantaneously.

The publishing of quarterly reports, changes in management or any news related to the reputation of a company may have drastic effects on the stock prices in the market (Mačaitytė & Virbašiūtė, 2018). The theory of Perfect Competition, which contains five criteria for market structure, states that buyers have access to all information such as price information and about the product from the past, present and future for a company and are able to make decisions based on this information (Stingler, 1957). With access to new information, buyers are able to determine appropriate prices for products sold in the market and make adjustments accordingly. With an accuracy rate of nearly 90 percent, using Twitter as a method to observe short term stock market movements could potentially make day trading more profitable for investment managers and individual stock traders around the world (Rao, 2014).

Herding Mentality

Asset prices fluctuate nearly every minute of everyday in an attempt to come to equilibrium, where the current value of the asset reflects the true value of the asset (Scarf, 1967). The standard asset model also known as the Capital Asset Pricing Model or CAPM (Black, Jensen, Scholes, 1972), assumes that markets are efficient (Reinganum, 1981) and that all market participants possess or have access to the same information (Brunnermeier, 2001). In reality, there is a wide divergence of information collected from market participants and how each individual interprets that information is subjective. There are almost always disagreements between analysts about the value of stocks, their projected growth or whether they are a buy, hold or sell for investors (Sadka & Scherbinam 2007).

For example, innovative American electric car manufacturer, Tesla, has sent analysts into a frenzy in the past with its volatile movements, drastic jumps and uncertain future. Between the months of August 2019 and February 2020, the stock leapt from $215 on August 26 to $901 on February 21, 2020, nearly a 320 percent increase. At its all-time high, analysts are still uncertain about its future performance because the company appeared weak on so many parameters of fundamental analysis, including a positive cash flow projection and previous earnings.

As of February 23, 2020, on CNNmoney.com, 12 out of 31 analysts recommend selling Tesla, 11 recommend to hold it, 2 believe it will underperform and the remaining 6 list it as a buy. In contrast, GovCapital.com lists it as a buy with 1 year projections reaching $1,187 and 5 year projections hitting $3,238. It is clear that with varying amounts of information available and differing interpretations of that information, asset prices and reflections of speculation in the market can cause stocks to fluctuate over a wide range of values.

Many of these varying opinions about future prices can be attributed to asymmetric information, the methods for calculating equilibrium prices as well as the way traders use this information when making buying and selling decisions. New information is constantly being released and "traders who do not receive a piece of new information are still conscious of the fact that the actions of other traders are driven by their information set. Therefore, uninformed traders can infer part of the other traders' information from the current movement of an asset's price" (Brunnermeier, 2001). Because of this lack of equal information exchange, it can be seen why some traders may latch onto the opinions of more prominent financial leaders or analysts and why phenomena such as herding and groupthink can occur often in the markets.

Groupthink is defined as the "psychological drive for consensus at any cost that suppresses dissent and appraisal of alternatives in cohesive decision making groups" (Janis, 1972). It is a psychological concept that has been widely renowned for it observed in many monumental events throughout history including the bay of pigs and the escalation of the Vietnam War (Janis, 1971). It often leads to poor decision making and outcomes that are not well thought out, but in the case of stocks, it can depend from which perspective you are looking.

For example, if you are currently in the market to buy shares of Apple and a well-respected investors such as Warren Buffett makes a public statement that it is a good stock to buy and he expects the price to increase, then millions of investors may be swayed by his statement even without doing any of their own research to back up his claims. "Agents with anticipatory preferences, linked through an interaction structure, choose how to interpret and recall public signals about future prospects. Wishful thinking (denial of bad news) is shown to be contagious when it is harmful to others, and self-limiting when it is beneficial"(Bénabou, 2012).

Because of the group action, the stock price will increase because there is a simulation of demand in the market (Goering, 1985). This would be great for investors that either own the stock or were planning to buy the stock at the same time as prices would rise. For investors that previously believed Apple was going to go down (Miller,

1977) before Buffett's claim, or were shorting the stock, then it would be a negative outcome for those investors. The phenomenon often occurs with social media. People are heavily influenced by the mentality of the masses even when it may not necessarily reflect their own ideas (Janis, 1972). In the case of Twitter, highly influential individuals including politicians, celebrities and CEOs will have an influence over the market just by submitting a single tweet about the economy, outlook on the future, or current trend they are obsessed with (Thomson, 2006).

There are also some scholars who argue social media platforms do not have drastic influence over stock performance because while people may have more access to information about markets; it is not necessarily accurate. Individuals can be more influenced by public opinion without doing their own critical and fundamental analysis before investing (Tirunillai & Tellis, 2012). Many well-known famous investors often recommend not getting into stocks in which the investor does not understand. With this idea of blindly following professional's recommendations, herd mentality, where inexperienced investors are willing to blindly follow the advice of experts without doing their own due-diligence on the investments, comes into play. This can be a dangerous investment method when approaching retirement age and using investments for pension funds (Bizzi & Labban, 2019).

An extension of group think is mob mentality. Mob mentality can be looked at as one of the reasons the impact of financial crises in the past have been so drastic (Kjaer, Teubner & Febbajo, 2011). For example, world renowned Korean researchers found "that herd behavior among all institutional investors increased significantly during the Korean economic crisis" (Kim Wei, 2002). Financial contagion is one reason why social media forecasting may only be good in the short term (Ranco, Aleksovski, Caldarelli, & Mozeti, 2011) as people can be easily swayed by public opinion causing them to overreact and act drastically.

With fear and panic during the crisis, the issues in the banking industry increased exponentially as large collecting of people looked to sell tanking assets or were left with overvalued real-estate with plummeting prices. Before the financial crisis of 2008, banks had few limitations regarding capital ratios, or the amount of money banks are required to hold compared to their risk weighted assets and large leeway with risk taking approaches (Furlong & Keeley, 1989). Now with more stringent regulations for liquidity and risk purposes, international organizations and governments are hoping to address these fundamental issues and thwart fears from the rapid spread of information through the public in order to prevent future crises of that scale.

Group think and herd mentality are heavily linked to the psychological approach to investing and behavioral finance. There are objections to both approaches to investing, the psychological approach and the fully rational approach. Some criticisms of the psychological approach include that “rational traders should arbitrage away mispricings [in the market] and that confused investors will learn their ways to good investing” while criticism for the rational approach are the opposite, stating that “irrational traders should arbitrage away efficient pricing and accurate investors will learn their way to bad decisions” (Hirshleifer, 2001). While both approaches have reasonable objections to their premises, it can be observed that investors more often act irrationally than rationally and have a greater influence on market returns.

Additionally, studies conducted by companies such as Deloitte looked at how customers can be influenced in their purchasing decisions by crowds. Peer reviews and public platforms are the new way to invest and having access to people’s opinions from around the world has proven to be heavily influential for investors. About “82 percent of US Internet consumers report being directly influenced by peer reviews in their purchasing decisions” (Chen, De, Hu & Hyoun Hwang, 2014).

It is easy for individuals to be heavily motivated by the opinions and thoughts of others (Zimbardo & Leippe, 1991). As we can see from the previous literature mentioned, investors today are looking for expert opinions through the use of social media. This enters into another area of study to look at; behavioral finance. Buyer trends, sentiment, and behaviors are all to be considered when looking at the effects of social media on markets. It is a way to look at patterns.

Many different forms of social interaction can be studied when looking into financial markets as nearly everything has an effect on them. From politics to natural disasters, many outside variables can influence investment outcomes (Kalampokis, Tambouris & Tarabanis, 2013). Other indicators can be predicted using social media as wel,l including box office revenues from new movie releases. From Twitter interactions and mentions, it has been possible for companies to anticipate how movies will do financially during their opening weekends (Asur, 2010). Twitter is a powerful tool that is able to be used in order to determine the relative success of new products and services with accuracy. From looking at this type of data and information, it is beneficial for Hollywood or other movie industries when launching or anticipating new films.

Isolated Examples

While there has been little research conducted on the topic of whether individual Twitter accounts can have an impact on the stock market, there are isolated cases where it is evident a comparison can be made. Many individuals value the opinions of celebrities and idolize their thoughts on popular products, movies and food. The age of influencers is a huge money making opportunity for those who have the following (Kapitan & Silvera, 2016). Those who have more than 100 thousand followers on Instagram are able to earn up to 2 thousand pounds per picture (approximately 2,500 USD), while those with between 2 million and 4 million followers can charge up to 13 thousand pounds (16,185 USD) (Lazazzera, 2018). It is estimated that U.S. reality star Kylie Jenner, a member of the famous Kardashian family, earns around 1.2 million dollars per post on Instagram (BBC, 2019). With such a heavy following on social media, it is easy to influence and market to others and have a real monetary impact on companies and earnings.

Celebrities with millions of followers can also have a large impact on the stock market with their content. Kylie Jenner, who made her mark as becoming the world's youngest self-made billionaire at age 21 in 2019 (Robehmed, 2019), now has her own line of cosmetics and clothing. In 2018, she tweeted about the relevance of snapchat after the launch of one of their updates asking, "sooo does anyone else not open Snapchat anymore? Or is it just me" (Levin, 2018) to her 24 million followers in mid February. Following the tweet, Snapchat stocks dropped by about 6 percent.

In 2013, well-known billionaire investor Carl Icahn was credited with raising Apple's valuation by 17.5 billion dollars with two tweets after writing about his large position in the company. His two tweets were used as a strategic long play in order to help drive Apple to spend some of their vast amounts of cash at the time in order to facilitate a stock repurchase (Metz, 2017). The first tweet stated, "We have a large position in Apple. We believe the stock to be extremely undervalued. Spoke to Tim Cook today. More to come." (Icahn, 2013) while the second said, "Had a nice discussion with Tim Cook today. Discussed my opinion that a larger buyback should be done now. We plan to speak again shortly." (Icahn, 2013). With these two tweets, Apple stock prices shot up over 5 percent to their highest value in 6 months to $489.57.

It is no secret that United States President, Donald Trump, is a huge proponent of Twitter. Due to the accessibility and popularity of the platform, it is a simple and effective way to get across his message to his followers. His tweets are known for being so erratic that some companies have created a new index to track market movements. American based investment bank J.P. Morgan has even devised its own index named the "Volfefe" index, combining the words volatility and Donald Trump's infamous misspelling of the word coverage in one of his tweets in 2017. The index looks specifically into tweets related to monetary policy and international trade, which could heavily affect their positions. In addition, Bank of America has reported to clients that markets tend to drop following days that Trump has tweeted more than usual (Stewart, 2019). While there are many examples of Trump's tweets impacting the market, there are a select few that have really caused drastic changes.

In 2016, Trump was able to wipe off 1 billion dollars from American multinational airplane, rocket and satellite manufacturer Boeing's valuation with his tweet in 2016 about the cost of the new Air Force one jet shortly after winning the US presidential election. He wrote, "Boeing is building a brand new 747 Air Force One for future presidents, but costs are out of control, more than $4 billion. Cancel order" (Trump, 2016). With a drop in stock price of 1 percent following the tweet, from $152.16 to $149.75, Boeing felt a major impact from Twitter right away (Revesz, 2016). Since 2016, Boeing stock has increased steadily, hitting a peak of $440 in March of 2019 demonstrating the short term effect of the tweet (Reddy, 2019).

The following year in 2017, Trump negatively tweeted about Japan based car manufacturer Toyota's decision to open a manufacturing plant in Guanajuato, Mexico, and shift their automobile production out of the United States. With his tweet threatening, "Build plant in US or pay big border tax!" (Trump, 2017) stock prices fell half a percent within minutes of the tweet's posting corresponding to a 1.2 billion dollar valuation loss for the car company (Revesz, 2017).

Previous events also indicate that individuals not in the direct spotlight can also have an impact on stock movements. In 2017, local Kansas reporter, Tim Hrencher, who was covering the merger between Kansas based utility company, Westar Energy and Missouri based holding company, Great Plains Energy, tweeted about the likelihood of a fall through of the proposed merger between the two companies. With only three favorites and one retweet it seems that this tweet would have little effect on the market, but following the release of this new information stating, "The attorney tells the KCC 'a merger makes sense -- just not this merger.'" He adds that this was proposed as a "take it or leave it proposal" (Hrencher, 2017). Prices for Westar Energy fell by 4.5 percent with a grand total of 7.7 percent down the following day (Comcowich, 2017). The merger did end up going through in May of 2018 and in late 2019 the two companies officially rebranded themselves as Evergy (Evergy, 2018). This case can be considered a bit more extreme than conventional observations and it can be assumed that few tweets with little to no recognition could have this drastic of an effect on a company.

From these isolated events, it can be seen that individuals such as celebrities and politicians and even local small town reporters with large and small followings are able to influence markets with their opinions. Twitter provides a real time platform to influence markets and for investors with a keen eye and quick reflexes, it may be possible to benefit off of these declarations. While hindsight is always 20/20, this paper’s research will look at the sentiment analysis and predictive power for tweets posted between the years 2016 and 2018, and will also look into the possibilities to anticipate future upcoming trends in the market.

Gap in the Literature:

While there has been extensive research conducted on the relationship between social media and the stock market, most often looking at Twitter, there still remains a gap in the literature looking at the specific effects that individuals can have on the stock market. The overarching question for research related to social media and the financial markets is whether or not personal Twitter accounts can have a direct influence over the actual outcomes in the market?

The results demonstrated throughout the relevant literature appear to be affirmative, although outcomes appear to vary based on the nature of tweets; if they are positive or negative. With large amounts of research being dedicated to the topic of social media and its effects on society, it is clear that there could be correlation between Twitter ‘mood’ and the movements in the stock market (Zhang, Fuehred & Gloor, 2011). This paper will continue to build off the previously conducted research in order to determine whether individual Twitter accounts can predict movements in the stock market with statistical significance, can reduce volatility in portfolios and achieve better returns.

Another aspect of the research to be addressed is exactly how to study the effects of social media. It can be difficult to quantify items relating to user sentiment and feelings because artificial intelligence (AI) is not yet sophisticated enough to identify human prose such as sarcasm. There is the numerical method of looking at the number of mentions and likes for posts, but the majority of conclusions are made from looking at the content or sentiment related to the situation. Qualitative vs quantitative analysis is critical for determining the effect of social media on the markets. Conventional news also has a different effect than social media. With an increasing amount of people getting their news from social media platforms, areas such as newspapers or television news stations are not having the same reach to users as they had in the past (Wenjing Du & Cao, 2013).

Other limitations lie in the way we access information. News sometimes takes hours or days to be released by prominent news stations because of liability issues or ensuring their sources are credible and by that time the stock movements up or down have already occurred. There is no overarching authority looking at the credibility of individuals' social media accounts. This lack of credibility has caused some issues in the past, mostly for Facebook, with the topic of “fake news” where fictitious information was being posed as legitimate and fact-checked information (Allcott & Gentzkow, 2017). On Twitter, people are able to interact with each other in an informal and natural way. Looking at Twitter also allows for instantaneous insight into nationwide and even global opinions and gives a leg up on the competition when investing in the short term.

The goal of this research will be to fill in some of the gaps in the literature as well as to observe links between using Twitter as a method to ensure more steady returns, lower volatility in portfolios and a larger profit for both individuals and firms.

Chapter 3: Methodology

The methodology used in this thesis is explanatory and will use the inductive method in order to test a new hypothesis. Due to the gap in the literature, it was determined the need for a study focused on the effects of individual Twitter accounts to be used as a short term stock market indicator for portfolio analysis as former studies have focused primarily on observing the social media platform as a whole. To fill in this gap, our study will focus on the relationship between the sentiments of the selected individuals, the number of users tweeting per trading day for the predetermined period, as well as the number of retweets. These variables will be compared to the stock market behavior of the S&P 500 index between the years of 2016 and 2018.

The methodology section of this research paper will be divided into two different sections. The first portion will cover the data and variable description. Topics are also inclusive of the reasoning behind the choice of variables to observe and the decision behind which individuals would be observed in the study as the data inputs.

The second portion of the methodology section will focus on the data extraction and the way it has been completed using data analysis tools such as the programming language of R. The section will culminate with an explanation of the statistical models used to interpret the collected data and methods for cleaning and organizing the data.

Data and Variable description

The dependent and independent variables are as outlined below.

The stock market behavior has been determined to be the dependent variable. The S&P 500 index was chosen as the market (Chen, Noronha & Singal, 2004) to be observed because it is the market that encompasses many of the largest companies in the United States of America. It has also been determined by the author to be more reflective of the overall market behaviors as it encompasses 500 of the largest companies by market capitalization. We expect similar movements in other markets including subdivisions such as the NASDAQ index and the Dow Jones Industrial Average. The S&P market includes most of the companies led by the CEOs selected for our study.

In order to study the behavior of the S&P market, the daily index adjusted close values between the years 2016 and 2018 were extracted and observed from Yahoo Finance. In total 753 trading days, days where the market was open and active, were selected for the purpose of the study. There are on average 252 investing trading days per year (Mayo, 2020).

The independent variables are comprised of the sentiments of the investors, retweets and users. One aspect of the data collected include the text of the tweets. Twitter currently allows 280 characters for their users, but in the first portion of the study until November 2017, Twitter's character limit was set at 140 characters. A higher character count enables users to give more complex thoughts and opinions. While this higher character limit may aid a bit with the classification of the tweets, it still does not completely erase the issues with the sentiment analysis ambiguity while deciphering the selected tweets.

It was necessary to determine a way to quantify the sentiment results of the tweets. We decided to utilize a binary system for classification in order to transform the data into an aggregatable form (Rajabi, Uzuner & Shehu, 2020). Every tweet was assigned a value of either p, o or n representing positive, neutral or negative sentiments respectively (Jiand, Yu, Zhou, Liu & Zhao, 2011). From the brief sentiment description, the sentiments were then converted into the numerical binary values of either +1, 0 or -1 corresponding to the positive, neutral and negative sentiments.

As mentioned in the literature review, sentiment analysis has been determined to be one of the most effective tools for Twitter/market analysis and was used as the foundation for the study (Prabowo & Thelwall, 2009). From there, overall daily sentiments of the users were determined finding the averages of the sentiments based on daily user numbers. We observed tweets from each trading day, found the sum of the binary values for each user per day and then used these individual results to aggregate an overall sentiment from the observed users. Completely positive tweets from all users would yield a result of 1.0 whereas entirely negative tweets yielded a -1.0 on the binary scale. Tweets tended to fall on the more positive spectrum for all users.

We also acknowledge that the science of sentiment analysis is not perfect and the possibility of misclassification of the data based on sentiments is high. It is likely not all of the recorded sentiment data reflects the true feelings of the user at that time due to ambiguity or lack of previous background information about current events (Bollen, Mao Zheng, 2011). It is also extremely difficult for current data science and machine learning tools to pick up on semantics in language such as sarcasm (Gonzales-Ibanez, Muresan & Washolder, 2011).

We also accept that this is an area of further study by increasing the complexity of the sentiment classification model to include more emotions and higher sophisticated technology. Some previous research had separated the sentiment into six classifications, including some emotions such as calmness and anxiety (Bollen, Mao Zheng, 2011). Additionally, emoticons displayed in tweets may not be accurately represented through sentiment analysis (Hu, Tang, Gao & Liu, 2013). Many users such as Ariana Grande, Katy Perry and Elon Musk often used emojis in their tweets making it more difficult to classify accurately.

Another independent variable that was observed is the number of retweets received on each Tweet. Retweets represent the amount of exposure an individual tweet has, as with a higher circulation rate; a higher number of people would see the Tweet on the platform expanding its influence on users on the platform (Pancer, Poole, 2016) and therefore would also achieve higher exposure to stock market participants that are also on the platform. Additionally, tweets sent out by famous or influential individuals often get recycled to other social media platforms or are used in conventional news stories, also increasing the exposure amount. The retweets are approximated using the number of recorded retweets for each individual tweet per user. Additionally, the number of retweets range between 0 and 1.6 million for the observed period over 3 years.

The last independent variable we observed is the number of users. From the data we were able to determine the number of users who tweeted on each trading day. The number of users range from 4 to 16 on any given day during the observed period with an overall average of 9 users tweeting per day. There were no recorded days where all users in the study tweeted.

Favorites for the tweets were chosen to be excluded in the study because of the high probability of multicollinearity, or high correlation, between the number of likes or favorites on a tweet and the number of retweets received.

Data Selection:

With millions of active users on Twitter from all over the world, ranging from those with tens of millions of followers to users with little to no followers, also known as bots, it can be difficult to determine which individuals would be able to have an influence over enough people to be able to be observed. For our research purposes, it was determined there are a few categories that individuals can fall into in order to have a large enough following to influence market decisions: politicians, CEOs/ company founders or executives, and celebrities such as actors, musicians and sports stars (Fraser & Brown, 2002).

For this research paper, it has been decided that the individuals to be observed would include ten of the most famous CEOs from some of the largest companies around the world and the ten most followed accounts on Twitter as of March 2020.

While only 20 personal accounts have been chosen to be observed, we were able to extract over 45,000 tweets over the three year period through the use of machine learning. We also think that these accounts are able to encompass a large number of the Twitter universe. Out of the 330 million active monthly users on Twitter, the total combined follower account for only these 20 accounts exceeds 1 billion. Of course, it can be assumed that many of the selected accounts have duplicated followers but with an average follower count in the tens of millions, it is evident of the scope of influence these individuals are able to have with their posts.

Businesses with large market capitalization have a significantly large impact on the market (Porter & Kramer, 2019). Therefore, it can be surmised that those in charge also have a great deal of influence and power over the companies with the ability to sway decision making processes. The list of chief executive officers included in the list are Amazon CEO Jeff Bezos, the richest man on the planet, Virgin founder Richard Branson and Tesla's founder and CEO Elon Musk.

Selected CEO Twitter Accounts:

Name	**Followers (2020)**
1. Bill Gates (Bill & Melinda Gates Foundation)	49.2 million
2. Elon Musk (Tesla)	32.7 million
3. Richard Branson (Virgin)	12.6 million
4. Tim Cook (Apple)	11.7 million
5. Sundar Pichai (Google)	2.6 million
6. Satya Nadella (Microsoft)	2.0 million
7. Warren Buffett (Berkshire Hathaway)	1.6 million
8. Jeff Bezos (Amazon)	1.3 million
9. Mark Zuckerburg (Facebook)	465 thousand
10. Brian Chesky (Airbnb)	313.1 thousand

In addition to CEOs, the other category will also include ten individuals with the largest following on Twitter as of March 2020. Those included are Portugese football star Cristiano Ronaldo, with 82.8 million followers, Canadian pop singer Justin Beiber with 110 million followers, and former forty fourth president of the United States of America, Barack Obama, currently the number one most followed account on Twitter with 113.3 million followers (Boyd, 2020).

Top 10 Most Followers Twitter Accounts:

Name	**Followers (2020)**
11. Barack Obama	113.3 million
12. Justin Beiber	110.0 million
13. Katy Perry	108.5 million
14. Rihanna	98.5 million
15. Taylor Swift	85.8 million
16. Cristiano Ronaldo	82.8 million
17. Lady Gaga	80.9 million
18. Ellen DeGeneres	79.6 million
19. Donald Trump	73.1 million
20. Ariana Grande	71.4 million

These celebrities, politicians and sport stars have been chosen for their widespread influence across the world. With fans and admirers all over, not only are they interested in what the celebrity is posting but users are also mentioning them through the attendance of concerts, sporting events or political gatherings. With over 70 million average followers per account, it can be seen from our data that a portion of the population is interested in hearing about the selected individual's thoughts and opinions. Our research and findings will focus on what effect their influence can have in the stock market.

Data Extraction

Twitter is fed with a constant stream of new information from its users every second of every day. With an estimation of over 500 million tweets sent per day or 6,000 per second (Twitter Usage Statistics, 2020), the only way to evaluate information and data of this size is to utilize data science tools. Due to the sheer number of tweets being posted every day, it is impossible to observe them in an accurate and efficient way without the use of technology and software in combination with excel (Laitenberger & Dreyer, 1998).

For this paper's research it has been decided to use the Twitter Application Programming Interface (API) named tweepy, a publically available open source development tool used in conjunction with the high-level programming language, Python, in order to collect and interpret the data in order to make the conclusions.

To begin, it was required to fill out a request for access to a Twitter developer account. As part of the request, it was also required to specify what the data will be used for, the type of data that will be collected such as individual usernames, the content, or sentiment analysis as this developer tool can be used for a number of different aspects including app development. After creating a developer account, access was granted to the public and private keys necessary for the tweepy package installation. We were able to download and install the Twitter related API package through Python, which was created to gather and analyze sentiment analysis. Through this package, developers are able to collect the specified information for which they are searching with the input of a few commands in the code. All the data collected was between the periods of 2016 to 2018 because of the number of influential political events during that time frame including the 2016 presidential election in the United States, the United Kingdom's decision to leave the European Union, also referred to as Brexit, and a number of radical political events sprinkled throughout the United States involving gun violence and racism.

All of the data collected is from primary sources, having come from the tweets themselves posted from individual users during the specified time period. There was no use of outside data from sources other than Twitter to influence the results of the data.

The data collected included the 19 out of the 20 individuals mentioned above (with the exception of Mark Zuckerburg, as no data was available), the date of their tweets, the number of likes, the number of retweets, any hashtags used, the wording of the tweet itself as well as the sentiment behind the tweet classified as either positive, negative or neutral. In total, approximately 45 thousand tweets were extracted and observed for the purpose of this study.

Following the data extraction, it was necessary to reorganize and clean the data. With the S&P 500 Index adjusted close numbers between the years of 2016 and 2018 (Yahoofinance.com), we determined the number of individuals from the selected 20 who had tweeted on that day. We also determined the overall sentiment for each trading day based on the sum of the binary scores per day. Following the individual sentiment collection and summarization, the result was divided by the number of users who tweeted that day, giving a sentiment score between negative and positive one. A score of negative one meant that all users who tweeted that day had only negatively classified tweets and positive one represented the opposite.

We excluded tweets that fell on weekends or holidays to observe the immediate impact of the tweets from the selected user on that day. With the short term nature of Twitter as mentioned in the literature review, those tweets just a few days old would not necessarily have an effect on the observed index. Therefore, only tweets posted on these 753 trading days were used as data inputs for our regression analysis.

With the tweets from non-trading days being excluded from our study, 36,139 observable tweets were left. In total, there were 4,635 negatively classified tweets, 19,342 neutral tweets and 12,162 positively identified tweets making up 12.8 percent, 53.5 percent and 33.65 percent, respectively. Neutral tweets were given a binary representation of 0 and were thought to have no influence on the independent variable. For final research purposes, 16,797 of the positively and negatively classified tweets were used in our statistical model over the three year time span.

Statistical Models

With the clean data, a statistical model needed to be applied in order to determine the trustworthiness of the results and test our proposal. For this paper, we chose to use two different statistical models, the ordinary least squares regression (OLS) model and the time series regression model, Auto-Regressive Integrated Moving Average (ARIMA) (Zhang, 2003). These models were determined to be the most appropriate for the data because they take into account all necessary factors that could have influenced the interpretation of the results (Atasalakis, Valavanua, 2009) including “seasonality, time and stationarity” (Cavalcante, Brasileiro, Souza, Nobrega & Oliviera, 2016). The models also enable users to work with exogenous variables, or variables outside of the model, to be adjusted when incorporated into it (Fair, 2013).

Ordinary Least Square Regression Equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Y = Dependent Variable

β_0 = Y Intercept

β_1 = Coefficient

X_1 = Independent Variable

ε = Random Error Term

For the study, the Ordinary Least Square (OLS) model has been decided to be used in order to control for the time element and the use of fixed effects on a yearly basis. By incorporating the fixed effects we are able to “remove omitted variable bias by tracking movements within groups across time, and are able to introduce dummy variables if needed for the missing or unknown characteristics (LaMotte, 1983). This model enables us to assume a linear relationship between the dependent variables of users, retweets and sentiments to the independent variable the S&P 500. In our paper, the data collection and design allow us to draw causal conclusions, thus the regression analysis for our data output is considered to be a causal analysis (Allison, 2014).

The OLS model is widely used as a way to measure impact between two variables in a linear regression equation (Hayes & Cai, 2007). It is also a way to minimize errors caused by the sum of squared errors, where there is a difference between observed and predicted variables. The OLS model is necessary to be included into our study alongside the ARIMA model because we are showing the impact of the independent variables on the dependent variable as opposed to using our model to forecast future outcomes in the market.

Assumptions associated with the use of the OLS model:

1. The linear regression is linear in parameters
2. The sample of observations is random
3. The conditional mean should be zero
4. There is no multicollinearity (Which was demonstrated through the use of the VIF test)
5. There is homoscedasticity and no autocorrelation. (Addressed with the use of the Breusch Pagan test)
6. Error terms should be normally distributed (Hayes & Cai, 2007)

ARIMA Model Equation:

$$Y_t = c + \emptyset_1 Y_{t-1} + \ldots + \emptyset_p Y_{t-p} + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} + e_t$$

Where:

- Y_t = the variable that will be explained in time t;
- c = constant or intercept;
- $\emptyset$ = coefficient of each parameter p;
- θ = coefficient of each parameter q;
- e_t = Residuals or errors in time t.

The ARIMA model is a type of time series regression most often used for forecasting purposes. Time series regressions are types of statistical models that are used in order to take data from data/ trends over time and train the model to make predictions about future movements (Hamilton, 1994). The regressions can be separated into three different types based on the data structure; time series, where observations of a variable are taken over time, cross-sectional data (Ostrom, 1990), where different variables are taken at the same point in time and finally pooled data (Sayrs, 1989), which is a combination of the previous two. Pooled data will be used for this research looking at the impact of the market from different variables over a three year time span, between the years 2016 and 2018.

Time series regression also relies on the foundation of stationarity, which demonstrates that the mean of the series remains constant over time (Nelson, 1990). Because the OLS model does not take into account stationarity as an input, the ARIMA model needed to be utilized in order to achieve proper regression outcomes.

While most data is not stationary, methods of differencing may also be used in order to de-trend the variables and make them stationary such as moving averages (Sutcliffe, 1994). The data collected for this study was not stationary and varied from the mean often for each variable. Therefore, as a way to de-trend the variables, moving averages spanning 20 trading days, representing average monthly movements, were used as one of the main components of the regressions.

As there are also trends for seasonality in the data set associated with holidays (Hu, 2013) and other major events occuring during the selected time period, it was also necessary to deseasonalize the data. In order to use the data in the appropriate statistical model, it was required to divide standardized variables (as explained below) by the previously calculated moving averages per observed trading day period.

As the scale of the independent variables are too drastic to compare, with retweets being in the millions per day, it was necessary to standardize the variables. Without the standardization of variables, the results would be misleading and unreliable (Frost, 2013).

Standardized Formula:

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean
$\sigma =$ Standard Deviation

We were able to standardize the chosen variables by taking the daily values, subtracting the average for the sample of 753 observations and then dividing the result by the standard deviation of the same sample. Following this recalculation, it was possible to analyze the regressions on the same y axis scale, making comparability easier between the variables.

The time series regression model was run in order to observe and justify each of the findings from the collected data in the study. It was necessary to run the model for each of the variables: the number of users, retweets and the average sentiment per day against the S&P 500 in order to observe the statistical significance and correlation between each of the variables.

With the trained model, predictions and forecasting can be utilized. Additionally, with extraneous variables with an outside force such as sentiments, the ARIMA model is relatively accurate in its predictions (Carta, Medda, Pili, Reforgiato & Saia, 2019). Although forecasting is outside of the scope of this study, our results from the use of the ARIMA model should give relatively accurate forecasts in short term daily movements of the S&P 500.

We also tested for potential biases and established the type of variance of the errors to identify any problems in our data, such as multicollinearity, and the types of standard errors that needed to be computed based on the variance of the errors and their potential autocorrelations. Multicollinearity is an error in data where independent variables are highly correlated and therefore can cause disturbances in regression results (Farrar & Glauber, 1967). As mentioned previously, likes or favorites had already been excluded from the study due to their high multicollinearity with the other independent variable, the number of retweets.

To test multicollinearity, the variance inflation factor (VIF) test (Craney & Surles, 2002) was carried out. We tested the relationships between all combinations of each of the dependent variables mentioned previously. The results show that the variance of the estimated coefficient of volatility is inflated by a factor of 1.03 and thus is very lowly correlated between any of the other variables. A VIF score under four is a significant value (Fox, 1997). As demonstrated below, the VIF of all of the other variables is low as well indicating extremely low correlation between each variable.

$$VIF_i = \frac{1}{1 - R_i^2}$$

VIF Test

Variable Statistic	**VIF Result**
Users	1.044
Sentiment	1.037
Retweets	1.036
S&P 500	1.057

To test for heteroscedasticity, the variance in the variables in the model (White, 1980), in the linear regression model to check whether the variance of the errors from the regression is dependent on the values of the independent variables, the Breusch Pagan test was also used (Breusch & Pagan, 1979) . The results show a very low p-value thus the null hypothesis of homoscedasticity is rejected and heteroskedasticity is assumed here. The results of the regression can be found in the following section.

After running the statistical tests and determining significance for our data, conclusions were able to be drawn from our results. The next chapter will detail the findings for this study as well as recommendations for future analysis.

Chapter 4: Findings

In this research paper, we investigated whether individual expression through the use of personal Twitter accounts are correlated to the movements of the S&P 500 index. The purpose of this research was to address the gap in the literature linking individual Twitter accounts and the stock market. In the past, studies had been focused on the predictive power of Twitter as a whole as opposed to particular, influential accounts. Additionally, we wanted to answer the question whether these selected Twitter accounts could be used as a short term stock market predictor for better portfolio analysis. The desired outcome for our research is to build on the previously conducted research linking Twitter and financial markets in order to ensure better investment strategies in the future.

Based on the results of the auto-regressive time series model and our data output, we have identified that "public mood state can indeed be tracked from the content of large-scale Twitter feeds by means of rather simple text processing techniques" (Bollen, Mao Zheng, 2011) by classifying tweets as either positive, negative or neutral as well as through the observation of retweet counts. From our results, we have detected that there is a link to stock market movements through the use of individual twitter accounts with relative accuracy.

The classification of the tweets are as detailed below. It can be seen that some users tweet much more often than others. For example, Richard Branson is extremely active on Twitter with over 9 thousand observed tweets over the three year period giving an average tweet count of 12.3 tweets per trading day. Comparatively, Jeff Bezos only had 127 observable tweets during the specified time period of our study with an average of 0.16 tweets per day. It is evident that the observed accounts may have biases in their representation in the study as well as the type of content they normally chose to express on their social media platforms (Liu & Zhang, 2012).

As it can be seen, many of the users tend to post much more positive tweets. The user with the highest number of negatively classified tweets is US president Donald Trump with just over 35 percent of his tweets for the observed period having a negative connotation. Looking at accounts with similar levels of daily activity for future studies could give a more accurate representation and reduce bias in the sentiment variable.

User	**Positive**	**Neutral**	**Negative**
Bill Gates	365	393	53
Richard Branson	3,153	5,528	605
Elon Musk	469	1,942	386
Ariana Grande	1,283	2,005	246
Brian Chesky	202	472	56
Rihanna	93	278	17
Lady Gaga	424	484	139
Justin Beiber	186	323	23
Katy Perry	389	1,105	153
Cristiano Ronaldo	349	315	18
Satya Nadella	133	102	18
Sundar Pichai	153	78	26
Taylor Swift	17	31	2
Jeff Bezos	65	58	4
Ellen Degeneres	2,135	3,360	245
Donald Trump	2,095	2,210	2,341
Tim Cook	196	65	55
Warren Buffett	157	183	32
Barack Obama	298	410	216
Total	**12,126**	**19,342**	**4,635**

As previously mentioned in the methodology, neutral tweets were awarded a binary score of zero and thought to not have a direct impact on the outcome of movements in the market. Therefore, only the positively and negatively classified tweets were used in the calculation for overall sentiment analysis in the regression model.

We chose to use the time series ARIMA model, it was able to be observed that through the selected variables including sentiment analysis, number of retweets and number of users, the movement of the S&P 500 can be forecasted with relative efficiency. From our results as an interpretation of the data collected, we determined that individual Twitter accounts had up to a 91.2 percent accuracy rating in predicting short term stock market movements. Former studies had achieved a prediction rate of just over 87 percent (Mao, Bollen, 2011).

The results of the time series regressions can be observed below. All variables were statistically significant, with the p value falling way below the standard .05 threshold, indicating significance at over 99 percent. Standardized results for each variable, as detailed in the methodology section, were used in lieu of the observed numerical observations in the regression models for comparison purposes.

The coefficients for the sentiment and retweet variables for the regression were positive as expected, but the coefficient for the user variable yielded a negative result. This was unexpected and did not support our original hypothesis which assumed a higher user count would also indicate a higher adjusted close for the stock index. From our results, we can only assume that less users tweeting on any given day may lead to an increase in the outcome of the S&P 500, while higher sentiments and retweet numbers would have the adverse effect leading to an increase in S&P 500 returns. These results leave an opportunity for future research to close this gap in the study and test additional hypotheses based on user count as the primary variable.

From the outcomes of the regression, it was determined that for every additional retweet there would be a positive impact on the S&P 500 by .00083. Relating to sentiment for each additional positively recorded tweet, there would be a positive impact of .0000928 on the market. Our regression results indicate that retweets actually have a greater positive impact on the market in comparison to sentiment analysis, which also contradicts the claims made in previous studies. Retweets indicate that users agree or show strong support for a specific tweet. Retweets are replicated onto another user's page whereas favorites or likes are only visible to that specific user. In previously conducted studies, it has been concluded that political preferences can even be predicted through only the use of retweets (Wong, Tan, Sen & Chiang, 2013).

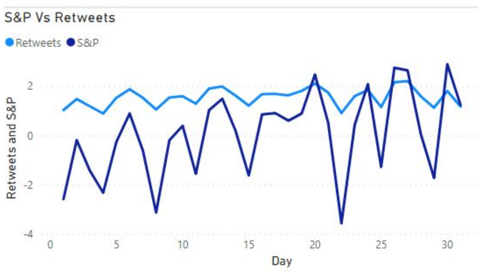
It was also reported that for each additional user in our study, there was a negative impact of .00054 on the market. In the literature there are studies focusing on the influence of users on Twitter but this was outside the scope of our particular study.

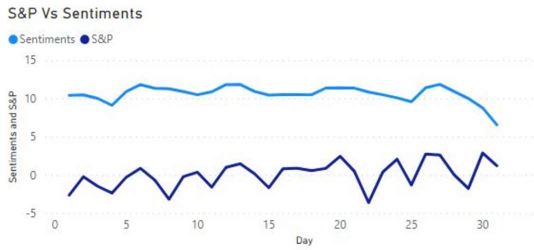
Regression Results

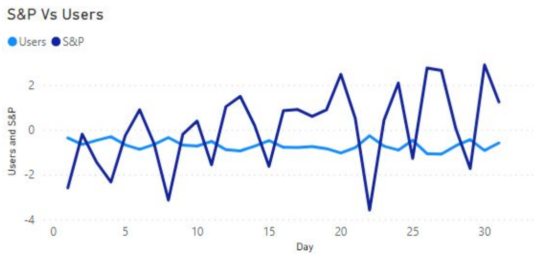
Variable	**Coefficient (Standardized) (Standard Error)**
Users	-.00054*** (.000116)
Sentiment	9.28 E -05*** (3.93E-05)
Retweets	.00083*** (.00028)
Observations	753
R^2	0.913
Adjusted R^2	0.912
F Statistic	7851.75*** (df = 3)
Note:	*p<0,1; **p<0.05; ***p<0,01

As it can be seen from the graphs below, the variables trend with the S&P 500 on a daily basis. We had observed 753 trading days over the three year time span chosen for the study, 2016-2018. All 45,000 observations were aggregated down as mentioned in the methodology section in order to match up with the S&P 500 trading day outputs. Tracking the daily movements of the independent variables against the dependent variable demonstrates a correlation in movements and predictive power. We decided to look at the relationships between the variables on a daily level as data collected represented a daily return.

The selected graphs are a subset of the data focusing on the overall trends of the variables focusing on the last month from 2018. From the three graphs below, a clear relationship can be observed between the dependent variable, the S&P 500 along with the independent variables, retweets and sentiments. The standardized variables were used for observational purposes. As mentioned previously, users had a negative correlation between the other variables and can be depicted below of having an inverse relationship between the S&P 500.







With the aforementioned results, the second portion of our research question was whether analysis from individual Twitter accounts could lead to better short-term portfolio management. Portfolio management is defined as a group of financial investments tailored to the risk profile and long term objectives of an individual, firm or institution (Cooper, Edgett & Kleinschmidt, 2001). There is a constant demand for higher returns, diversification of risk and ability to maintain these strategies for long periods of time.

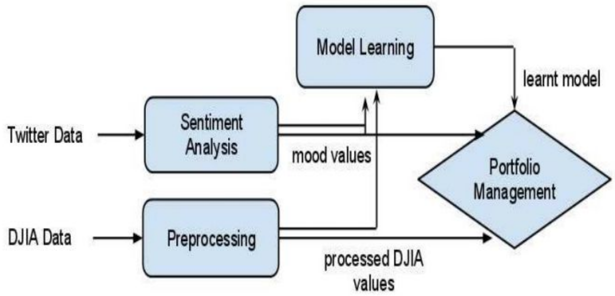
Based on our previously mentioned outcomes, we looked to test the hypothesis of whether the results of sentiment analysis and retweets could be used to achieve higher returns than conventional fundamental analysis. From earlier research, there have been relationships established between sentiment analysis and higher portfolio management results. Because we were able to establish a statistically significant result between sentiment and stock market movements without research, we can assume similar results as the former literature.

In the paper titled, “News impact on stock price return via sentiment analysis”, published in 2014, similar sentiment classifications of positive, negative and neutral were used for methodology purposes. The authors were able to establish a link between sentiment and stock price returns in the market. Their conclusions came to a correlation mean result of over 67 percent (Li, Xie Chen, Wang & Deng, 2014). While this is not extremely high correlation, their results indicated a two-thirds chance of predicting stock movements.

As mentioned in the introduction section of this paper, in another journal published in 2015, there was a study conducted where traders using social media for their investments averaged “2.07 percent better returns compared to solely using the historical model” (Nguyen & Shirai, 2015). A two percent better return is massive when speaking about hedge funds and portfolio managers who are in charge of funds with hundreds of millions of dollars being invested.

There are various other research studies that have also made links between portfolio performance using twitter sentiment results along with market sentiments for equity stocks in the past. Authors Anshul Mittal and Arpit Goel from Stanford University, in the paper titled, “Stock Prediction Using Twitter Sentiment Analysis” published in 2012, came to similar conclusions when utilizing Self Organizing Fuzzy Neural Networks (SOFNN) in addition to machine learning and sentiment analysis tools on the Dow Jones Industrial Average. Their findings concluded that they were able to predict stock market movements with just over 75 percent accuracy (Mittal, Goel, 2012).

As outlined in their paper, our process follows a similar flow from data extraction to the use of model learning techniques and concluding with the observation of better portfolio management possibilities utilizing the results of the study.



(Mittal, Goel, 2012)

With future technological advances making it easier to extract data from Twitter, this collective information will be extremely beneficial to use for asset and portfolio managers while making important decisions about whether to buy or sell their stocks. Investors will be happier with higher returns as well as companies and governments being more content with the mitigation of risk.

With accuracy rates over 65 percent from previous studies as well as 90 percent from this study, the results demonstrate a strong indication of how stocks will move in the near future. Although the results obtained for these studies do not give an indication of how Twitter sentiment analysis can influence stocks in the long term, this could be an additional area of study for future research in upcoming years.

Summary

The use of social media has exploded over the past two decades throughout the world. With users from every country across the globe and representation from all social classes, social media gives an in-depth view of real time opinions and enables a larger platform base for users to display thoughts and opinions (Treem & Leonardi, 2013) outside of conventional news sources. Although there may still be biases in social media (Lin, Bagrow & Lazer, 2011), users specifically focusing on Twitter, are able to give their feelings and opinions about the current events happening in the media and their lives at that moment.

It also can be seen the platforms' influence on societal behavior and its link to effects in capital markets. Through the use of sentiment analysis and retweet counts, a relationship can be established between observed variables and movements in the market. More retweets indicate a positive movement in the S&P 500 as well as positively conveyed sentiments on the platform. Retweets are often used in support of a message on Twitter and have a more supportive connotation (Wong, Tan, Sen & Chiang, 2013). Higher number of retweets indicate a high level of agreement and backing from users, more so than other variables such as favorites would have. We have also come to the conclusion that users have the opposite result, where the greater number of users have a negative impact on the market when observing individual Twitter accounts.

From the previously established link between Twitter mood and stock market movements, this research paper looks to open the conversation about quantifying the impact individuals can have on society and their influence on financial markets. While we cannot come to the conclusion individuals have a direct impact on the movements of the market, we can witness through our results a link between positive and negative movements in the S&P 500 index through the use of sentiment analysis and public response through the observations of retweets posted from individuals.

In the past, researchers had used millions of Twitter accounts to observe the effects of the social media platform as a whole (Bollen, Mao Zeng, 2011). As mentioned in the literature review, using Twitter in its entirety follows the notion of the "wisdom of the crowd," where the average of all observations is more close to the actual result in comparison to anything observed from a given individual (Suroweicki, 2004). While we did not operate our hypothesis under the same theory, from our observations, we came to nearly identical conclusions as previously conducted research with the data collected from just19 accounts.

Therefore, it can be seen that influential people such as CEOs, politicians and celebrities can have a direct influence over others, which in turn has an influence over those peoples' feelings and investment behavior. Psychology plays a huge role in the investment world (Shefrin, 2002) whether it is realized or not. From the outcomes of this study, we can see that emotions and feelings expressed through the use of social media can lead to monetary gains and losses if one chooses to invest in the stock market.

Finally, based on previous literature we are able to come to the conclusion that better portfolio analysis can come from the observation of targeted personal Twitter accounts. In the past, there have been higher returns and increased performance based on investment strategies relating to sentiment analysis. Operating under the same principle, it can be assumed that new investment strategies based on this paper's results would have similar outcomes. With less work and observations, those utilizing sentiment analysis and retweet counts could earn higher returns.

Conclusions and Managerial Implications

With the demand for portfolio and hedge fund managers to have constant and higher returns, there is always a need for new ways to mitigate risk. As mentioned previously, high risk is often associated with high reward (Baskerville, 2001). If portfolio managers have not been achieving expected returns with their investments over a certain amount of time, they may take more on more aggressive positions and more risk. It may have a beneficial payoff, but also encompasses the chance of detrimental losses.

Lack of excessive risk protocols has led to recessions in the past, most recently the 2007 Financial Crisis (Kirkpatrick, 2009). This has caused some managers to look "for opportunities in lower rated, less liquid, off-benchmark bonds, adding higher credit and liquidity risk (potentially higher currency risk)" (What Price Yield?, 2015). The European Securities and Market Authorities (ESMA) has begun to establish guidelines regulating the remuneration of fund managers of alternative funds, including hedge funds, private equity funds and real estate funds. The goal is aimed at cracking down on organizational structures which may encourage excessive risk taking (ESMA, 2013).

While the world is moving towards more advanced regulatory policies regarding risk management such as those that have begun to happen in Europe and the United States, there have been increasing pressures to instill stricter policies across the board in the financial markets. After the financial crisis, there have been movements to crack down on excessive risk taking and limiting risky practices for asset managers and large financial firms that were previously deemed "too big to fail" (Stern & Feldman, 2004)

From this research, as an extension of former research on the topic of the relationship between Twitter and the stock market, it can be concluded that individual Twitter accounts can be used as a short-term stock market predictor. Based on this information, there are many use cases for this type of forecasting and beneficial managerial implications. With better and more accurate predicting power, it can serve as a way to reduce risk.

There is no way to tell the future, but through the use of Twitter sentiment analysis and forecasting, it could be a way for fund managers to craft better portfolios with higher returns. With a higher confidence in stock movements, less risky positions would be needed to achieve similar results leading to more satisfied upper management as well as shareholders.

This research primarily focused on the returns for the S&P 500 and it is not exhaustive. While it does not take into account individual security movements or returns, companies that tend to have a high correlation with the market, for example, Microsoft, with a five year beta around 1 (YahooFinance, 2020), the results of this study can be implemented into management techniques and approaches.

Limitations and Further Research

Limitations for this research are as outlined in the following paragraphs.

By focusing research on a few select influential people, it is more manageable to collect and interpret data on an individual level as opposed to looking at every single account on the social media platform in its entirety. However, the used sample size is considered one limitation for this study.We would like to acknowledge the fact that the user count of 20 individuals selected is a limitation of the study and would look to include a larger sample size for future research.

The second limitation for this study is the availability of data for all accounts As we were unable to collect data from all of the selected accounts, it reduced our study to only 19 users as opposed to 20. Even with a selected sample size of 20, we would hope to increase the number of observations for the future. There are opportunities for additional research using a larger number of observations to improve statistical accuracy and gain a wider inside into the scope of influence of individuals.

A third limitation includes the sentiment analysis results. As mentioned previously in this paper, artificial intelligence is not yet capable of detecting social nuances or sarcasm for accurate readings. There are also biases associated with the sentiment analysis of the tweets. Using a system of classification limited to only positive, negative and neutral could also be expanded on to include more emotions as many tweets are ambiguous or contain contrasting feelings making it difficult to classify. With further developments in artificial intelligence technology, results can be more accurate and significant. The use of emoticons across the users could also have had an influence on the results as classification could have been incorrect or misleading.

Finally, the statistical method for this research could be another potential limitation. As it was decided to use the auto-regressive integrated moving average regression model for the data, we acknowledge the fact that there are other possible statistical models that could also be used for further research. We chose to use the auto-regressive integrated time series model because it addressed the necessary factors and enabled easily comparable data structures for our study.

For example, RF Engle's Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) statistical model, an extension of his AutoRegressive Conditional Heteroskedasticity (ARCH) model (Engle, 1982) could have been another viable statistical option for the interpretation of our results. "GARCH is used extensively within the financial industry as many asset prices are conditional heteroskedastic" '(Auquan, 2017). Heteroskedasticy refers to the variance in the variables in the model (White, 1980) and the concept exists in finance due to the volatile nature of asset returns. With more extensive statistical background and abilities, enhancement of the model used could be an opportunity for future research.

In conclusion, this paper has begun to answer the question of whether personal Twitter accounts can be used as a stock market indicator. Our results looked at the impact of the independent variables. With further improvements and methods to address the limitations in the study, we hope further research will be conducted to build off of our findings in order to create forecasts and aid investors in their decision making processes.

References

About. (n.d.). Retrieved from https://about.twitter.com/en_gb.html

Ahmad, Z. (2020). Short-run overreaction, stock prices and investors' irrationality in the Kuala Lumpur Stock Exchange. International Journal of Management Studies, 11(2), 1-19.

Ajour El Zein, S., Consolacion-Segura, C., & Huertas-Garcia, R. (2020). The Role of Sustainability in Brand Equity Value in the Financial Sector. Sustainability, 12(1), 254.

Alaghi, K. (2012). Asset and systematic risk. African Journal of Business Management, 6(9), 3504.

Alexander, S. S. (1961). Price movements in speculative markets: Trends or random walks. Industrial Management Review (pre-1986), 2(2), 7.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-36.

Allison, P. (2014). Prediction vs. causation in regression analysis. Statistical Horizons.

Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01 (pp. 492-499). IEEE Computer Society.

Atsalakis, G.S.; Valavanis, K.P. Surveying stock market forecasting techniques—Part II: Soft computing methods. Expert Syst. Appl. 2009, 36, 5932–5941

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. International Journal of Forecasting, 36(2), 334-357.

Auquan. (2017, December 19). Time Series Analysis for Financial Data VI- GARCH model and predicting SPX returns. Retrieved April 18, 2020, from https://medium.com/auquan/time-series-analysis-for-finance-arch-garch-models-822f87f1d755

Baker, D. (2008). The housing bubble and the financial crisis. Real-world economics review, 46(20), 73-81.

Baker, H. K., Nofsinger, J. R., & Puttonen, V. (2020). Common Investing Pitfalls that Can Separate You from Financial Security and Success. In The Savvy Investor's Guide to Avoiding Pitfalls, Frauds, and Scams. Emerald Publishing Limited.

Baker, S. R., Bloom, N., Davis, S. J., Kost, K. J., Sammon, M. C., & Viratyosin, T. (2020). The unprecedented stock market impact of COVID-19 (No. w26945). National Bureau of Economic Research.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, February). Everyone's an influencer: quantifying influence on twitter. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 65-74).

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. Handbook of the Economics of Finance, 1, 1053-1128.

Baskerville, R. (2001). Conducting action research: High risk and high reward in theory and practice. In Qualitative research in IS: Issues and trends (pp. 192-217). IGI Global.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. Review of general psychology, 5(4), 323-370.

Bénabou, R. (2012). Groupthink: Collective delusions in organizations and markets. Review of Economic Studies, 80(2), 429-462.

Benabou, R., & Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. The Quarterly Journal of Economics, 107(3), 921-958.

Bender, R. (2013). Corporate financial strategy. Routledge.

Bizzi, L., & Labban, A. (2019). The double-edged impact of social media on online trading: Opportunities, threats, and recommendations for organizations. Business Horizons, 62(4), 509-519.

Black, F., Jensen, M. C., & Scholes, M. (1972). The capital asset pricing model: Some empirical tests. Studies in the theory of capital markets, 81(3), 79-121.

Blomstrom, M., Kokko, A., & Zejan, M. (2000). Foreign direct investment: Firm and host country strategies. Springer

Bodie, Z. K. A. & Marcus, J.(2004): Essential of Investments.

Bodie, Z., Kane, A., & Marcus, A. J. (2003). Essentials of Investments. International Edition.

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

Bower, J. L. (1972). Managing the resource allocation process: A study of corporate planning and investment. Homewood, IL: Irwin

Boyd, J. (2020, February 28). The 20 Most Followed Accounts on Twitter. Retrieved March 3, 2020, from https://www.brandwatch.com/blog/most-twitter-followers/

Brockwell, P. J., & Davis, R. A. (1991). Time Series: Theory and Methods (2nd ed.). New York: Springer-Verlag.

Brown, S. J. (2020). The Efficient Market Hypothesis, the Financial Analysts Journal, and the Professional Status of Investment Management. Financial Analysts Journal, 1.

Brunnermeier, M. K., Markus, K., & Brunnermeier, M. K. (2001). Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding. Oxford University Press on Demand.

Cardona, O. D. (2013). The need for rethinking the concepts of vulnerability and risk from a holistic perspective: a necessary review and criticism for effective risk management. In Mapping vulnerability (pp. 56-70). Routledge

Carta, S., Medda, A., Pili, A., Reforgiato Recupero, D., & Saia, R. (2019). Forecasting E-Commerce Products Prices by Combining an Autoregressive Integrated Moving Average (ARIMA) Model and Google Trends Data. Future Internet, 11(1), 5.

Cavalcante, R.C.; Brasileiro, R.C.; Souza, V.L.; Nobrega, J.P.; Oliveira, A.L. Computational Intelligence and Financial Markets: A Survey and Future Directions. Expert Syst. Appl. 2016, 55, 194–211.

Chan, L. K., & Lakonishok, J. (2004). Value and growth investing: Review and update. Financial Analysts Journal, 60(1), 71-86.

Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2013). Customers as advisors: The role of social media in financial markets. Working paper.

Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. The Review of Financial Studies, 27(5), 1367-1403.

Chen, H., Noronha, G., & Singal, V. (2004). The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation. The Journal of Finance, 59(4), 1901-1930.

Chen, R., & Lazer, M. (2013). Sentiment analysis of twitter feeds for the prediction of stock market movement. stanford edu Retrieved January, 25, 2013.

Chong, T. T. L., & Li, X. (2019). Understanding the China–US trade war: causes, economic impact, and the worst-case scenario. Economic and Political Studies, 7(2), 185-202.

Clement, J. (2019, November 20). Global digital population 2019. Retrieved January 9, 2020, from https://www.statista.com/statistics/617136/digital-population-worldwide/.

Comcowich, W. C. W. J. (2017, May 9). Monitoring Twitter: The New Mandate for Financial Professionals and Investor Relations. Retrieved February 9, 2020, from https://glean.info/monitoring-twitter-new-mandate-financial-professionals-investor-relations/

Cooper, R. G., Edgett, S. J., & Kleinschmidt, E. J. (2001). Portfolio management for new products.

CNNMoney. (2020, February 21). TSLA. Retrieved February 23, 2020, from https://money.cnn.com/quote/forecast/forecast.html?symb=tsla

Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. Quality Engineering, 14(3), 391-403.

Cromwell, J. B., Hannan, M. J., Labys, W. C., & Terraza, M. (1994). Multivariate tests for time series models. Thousand Oaks, CA: Sage Publications.

Cromwell, J. B., Labys, W. C., & Terraza, M. (1994). Univariate tests for time series models. Thousand Oaks, CA: Sage Publications.

Crosbie, J., & Sharpley, C. F. (1989). DMITSA: A simplified interrupted time-series analysis program. Behavior Research Methods, Instruments & Computers, 21(6), 639-642.

Cummings, W. (2020, January 23). Trump sets record for most tweets in a single day since he took office. Retrieved March 17, 2020, from https://www.usatoday.com/story/news/politics/2020/01/23/trump-record-most-tweets-since-taking-office/4551815002/

Dang, H. V., & Lin, M. (2016). Herd mentality in the stock market: On the role of idiosyncratic participants with heterogeneous information. International Review of Financial Analysis, 48, 247-260.

De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact?. The Journal of finance, 40(3), 793-805.

DeLong, J. B., A. Shleifer, L. Summers, and R. J. Waldmann, 1990a, Noise trader risk in financial markets, Journal of Political Economy 98, 703–738.

DeLong, J. B., A. Shleifer, L. Summers, and R. J. Waldmann, 1990b, Positive feedback investment strategies and destabilizing rational speculation, Journal of Finance 45, 375–395.

DeLong, J. B., A. Shleifer, L. Summers, and R. J. Waldmann, 1991, The survival of noise traders in financial markets, Journal of Business 64, 1–20.

Deo, J. (2019, November 6). How hard is technical analysis. Retrieved March 17, 2020, from https://www.johndeoresearch.com/how-hard-is-technical-analysis/

Dijkmans, C., Kerkhof, P., & Beukeboom, C. (2020). Adapting to an Emerging Social Media Landscape: The Rise of Informalization of Company Communication in Tourism. In Information and Communication Technologies in Tourism 2020 (pp. 3-14). Springer, Cham.

Dodd, R. (2012). Markets: Exchange or over-the-counter. Finance & Development.

Duffie, D., Gârleanu, N., & Pedersen, L. H. (2005). Over‐the‐counter markets. Econometrica, 73(6), 1815-1847.

Dybvig, P. H., & Ross, S. A. (2008). Arbitrage. The New Palgrave Dictionary of Economics: Volume 1–8, 188-197.

Eades, K. M., Hess, P. J., & Kim, E. H. (1985). Market rationality and dividend announcements. Journal of Financial Economics, 14(4), 581-604.

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. The Journal of Finance, 66(1), 67-97.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007

Engle, R. (2004). Risk and volatility: Econometric models and financial practice. American economic review, 94(3), 405-420.

ESMA. (2013, February). Regulation - ESMA rules aim to curb excessive risk taking by alternative fund managers. Retrieved April 21, 2020, from https://www.next-finance.net/ESMA-rules-aim-to-curb-excessive

Evergy. (2018, May 24). Merger of Great Plains Energy and Westar Energy Approved, Paving Way for Benefits to Customers, Communities and Shareholders. Retrieved March 2, 2020, from http://www.greatplainsenergy.com/news-releases/news-release-details/merger-great-plains-energy-and-westar-energy-approved-paving-way

Fabozzi, F. J. (Ed.). (2008). Handbook of Finance, Financial Markets and Instruments (Vol. 1). John Wiley & Sons.

Fair, R. C. (2013). Macroeconometric modeling. Unpublished manuscript, November, 11.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. The Review of Economic and Statistics, 92-107.

Fender, J. (2020). Beyond the efficient markets hypothesis: Towards a new paradigm. Bulletin of Economic Research.

Flannery, M. J. (2017). Stabilizing large financial institutions with contingent capital certificates. In The Most Important Concepts in Finance. Edward Elgar Publishing.

Fox, J. (1997). Applied regression analysis, linear models, and related methods. Sage Publications, Inc.

Fraser, B. P., & Brown, W. J. (2002). Media, celebrities, and social influence: Identification with Elvis Presley. Mass Communication & Society, 5(2), 183-206.

Froot, K. A., Scharfstein, D. S., & Stein, J. C. (1992). Herd on the street: Informational inefficiencies in a market with short-term speculation. The Journal of Finance, 47(4), 1461-1484.

Froot, K. A., Scharfstein, D. S., & Stein, J. C. (1993). Risk management: Coordinating corporate investment and financing policies. the Journal of Finance, 48(5), 1629-1658.

Frost, J. (2013). Regression analysis: How do I interpret R-squared and assess the goodness-of-fit. The Minitab Blog, 30.

Gallistel, C. R. (1992). Classical conditioning as a nonstationary, multivariate time series analysis: A spreadsheet model. Behavior Research Methods, Instruments & Computers, 24(2), 340-351.

Garber, P. M. (1996). Managing risks to financial markets from volatile capital flows: the role of prudential regulation. International Journal of Finance & Economics, 1(3), 183-195.

Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., Schoen, H., Gloor, P., ... & Tarabanis, K. (2013). Understanding the predictive power of social media. Internet Research.

Gidofalvi, G., & Elkan, C. (2001). Using news articles to predict stock price movements. Department of Computer Science and Engineering, University of California, San Diego

Goering, P. A. (1985). Effects of product trial on consumer expectations, demand, and prices. Journal of Consumer Research, 12(1), 74-82.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2012). Assessing the bias in communication networks sampled from twitter. arXiv preprint arXiv:1212.1684.

González-Ibánez, R., Muresan, S., & Wacholder, N. (2011, June). Identifying sarcasm in Twitter: a closer look. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2 (pp. 581-586). Association for Computational Linguistics.

Gov Capital. (2020, January 6). Tesla (TSLA) stock price prediction is 1187.880 USD. The Tesla stock forecast is 1187.880 USD for 2021 February 23, Tuesday; and 3238.623 USD for 2025 February 23, Sunday. Retrieved February 23, 2020, from https://gov.capital/stock/tsla-stock/

Greenwald, B. C., & Stiglitz, J. E. (1993). Financial market imperfections and business cycles. The Quarterly Journal of Economics, 108(1), 77-114.

Griffith, J., Najand, M., & Shen, J. (2020). Emotions in the stock market. Journal of Behavioral Finance, 21(1), 42-56.

Gross, J. J. (2013). Emotion regulation: taking stock and moving forward. Emotion, 13(3), 359.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. Multivariate Behavioral Research, 40(2), 207-233.

Hamilton, J. D. (1994). Time series analysis (Vol. 2, pp. 690-696). New Jersey: Princeton.

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior research methods, 39(4), 709-722.

He, W., Guo, L., Shen, J., & Akula, V. (2016). Social media-based forecasting: A case study of tweets and stock prices in the financial services industry. Journal of Organizational and End User Computing (JOEUC), 28(2), 74-91.

Hendershott, T., & Moulton, P. C. (2011). Automation, speed, and stock market quality: The NYSE's hybrid. Journal of Financial Markets, 14(4), 568-604.

Hirshleifer, D. (2001). Investor psychology and asset pricing. The Journal of Finance, 56(4), 1533-1597.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. American economic review, 92(5), 1644-1655.

How much does Kylie Jenner earn on Instagram? - CBBC Newsround. (2019, July 26). Retrieved February 9, 2020, from https://www.bbc.co.uk/newsround/49124484

Hughes, N. C. (2005). A trade war with China?. Foreign Affairs, 94-106.

Hu, W. (2013). Real-time Twitter sentiment toward thanksgiving and Christmas holidays.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013, May). Unsupervised sentiment analysis with emotional signals. In Proceedings of the 22nd international conference on World Wide Web (pp. 607-618).

Hunt, A., Moyer, S., & Shevlin, T. (2000). Earnings volatility, earnings management, and equity value. Unpublished working paper. University of Washington.

Janis, I. L. (1972). Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (pp. 56-65). ACM.

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011, June). Target-dependent twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 151-160). Association for Computational Linguistics.

Kahn, M. J., & Baum, N. (2020). Basic Personal Finance and Investing. In The Business Basics of Building and Managing a Healthcare Practice (pp. 47-52). Springer, Cham

Kapitan, S., & Silvera, D. H. (2016). From digital media influencers to celebrity endorsers: attributions drive endorser effectiveness. Marketing Letters, 27(3), 553-567.

Kashif, M., Shaikh, A., & Rehman, M. U. (2020). Are financial investors prone to exogenous (cricket) sentiments that affects equity investment decision and induces volatility in stock market?. Estudios de economía aplicada, 38(2), 8.

Kellerman, A. (1985). The evolution of service economies: a geographical perspective. The Professional Geographer, 37(2), 133-143.

Kim, W., & Wei, S. J. (2002). Foreign portfolio investors before and during a crisis. Journal of International Economics, 56(1), 77-96.

Kirkpatrick, G. (2009). The corporate governance lessons from the financial crisis. OECD Journal: Financial Market Trends, 2009(1), 61-87.

Kjaer, P. F., Teubner, G., & Febbrajo, A. (Eds.). (2011). The financial crisis in constitutional perspective: The dark side of functional differentiation. Bloomsbury Publishing.

Kollewe, J. (2011, October 6). Apple stock price falls on news of Steve Jobs's death. Retrieved December 29, 2019, from https://www.theguardian.com/technology/2011/oct/06/apple-stock-steve-jobs.

Kothari, S. P. (2001). Capital markets research in accounting. Journal of accounting and economics, 31(1-3), 105-231.

Kouloumpis, E., Wilson, T., & Moore, J. (2011, July). Twitter sentiment analysis: The good the bad and the omg!. In Fifth International AAAI conference on weblogs and social media.

Laitenberger, O., & Dreyer, H. M. (1998, November). Evaluating the usefulness and the ease of use of a web-based inspection data collection tool. In Proceedings Fifth International Software Metrics Symposium. Metrics (Cat. No. 98TB100262) (pp. 122-132). IEEE.

Lakonishok, J., & Shapiro, A. C. (1986). Systematic risk, total risk and size as determinants of stock market returns. Journal of Banking & Finance, 10(1), 115-132.

LaMotte, L. R. (1983). Fixed-, random-, and mixed-effects models. In Encyclopedia of Statistical Sciences, ed. S. Kotz, N.

Lazazzera, M. (2018, January 14). How Instagram influencers turn followers into dollars. Retrieved February 9, 2020, from https://www.ft.com/content/fc964254-155f-11e7-b0c1-37e417ee6c76

Lee, R. S. (2020). Introduction to Quantum Finance. In Quantum Finance (pp. 3-16). Springer, Singapore.

Levin, A. (2020). Influence Is an Outcome, Not a Profession. In Influencer Marketing for Brands (pp. 19-46). Apress, Berkeley, CA.

Levin, S. (2018, February 24). 'We're watching a company explode': is Snapchat becoming irrelevant? Retrieved January 10, 2020, from https://www.theguardian.com/technology/2018/feb/23/snapchat-redesign-scandal-kylie-jenner-what-went-wrong.

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. Knowledge-Based Systems, 69, 14-23

Li, Y. (2019, August 14). Dow tanks 800 points in worst day of 2019 after bond market sends recession warning. Retrieved February 23, 2020, from https://www.cnbc.com/2019/08/14/stock-markets-wall-street-in-focus-amid-earnings-economic-data.html

Lim, K. W., & Buntine, W. (2014, November). Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In Proceedings of the 23rd ACM international conference on conference on information and knowledge management (pp. 1319-1328).

Lin, Y. R., Bagrow, J. P., & Lazer, D. (2011, July). More voices than ever? quantifying media bias in networks. In Fifth International AAAI Conference on Weblogs and Social Media.

Lin, Y. (2019, November 30). 10 Twitter Statistics Every Marketer Should Know in 2020. Retrieved December 26, 2019, from https://www.oberlo.com/blog/twitter-statistics.

Lintner, J. (1975). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In Stochastic optimization models in finance (pp. 131-155). Academic Press.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Liu, B., Xia, X., & Xiao, W. (2020). Public information content and market information efficiency: A comparison between China and the US. China Economic Review, 101405.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining text data (pp. 415-463). Springer, Boston, MA.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. Advances in experimental social psychology, 20(3), 255-295.

Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. Information Systems Research, 24(1), 146-163.

Mačaitytė, I., & Virbašiūtė, G. (2018). Volkswagen emission scandal and corporate social responsibility–a case study.

Madura, J. (2020). Financial Markets & Institutions. Cengage learning.

Mainelli, M. (2004). Personalities of risk/reward: Human factors of risk/reward and culture. Journal of Financial Regulation and Compliance, 12(4), 340-350.

Makazhanov, A., Rafiei, D., & Waqar, M. (2014). Predicting political preference of Twitter users. Social Network Analysis and Mining, 4(1), 193.

Malkiel, B. G. (1989). Is the stock market efficient?. Science, 243(4896), 1313-1318.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383-417.

Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, twitter and search engine data. arXiv preprint arXiv:1112.1051.

Mao, Y., Wei, W., Wang, B., & Liu, B. (2012, August). Correlating S&P 500 stocks with Twitter data. In Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research (pp. 69-72).

Martin, K. (2019, February 7). The World's Most Important Stock Exchanges - Admiral Markets. Retrieved April 13, 2020, from https://admiralmarkets.com/education/articles/general-trading/worlds-most-important-stock-exchanges

Mayo, H. B. (2020). Investments: an introduction. Cengage Learning.

McDonald, R., & Siegel, D. (1986). The value of waiting to invest. The quarterly journal of economics, 101(4), 707-727.

McGregor, S. C., & Molyneux, L. (2020). Twitter's influence on news judgment: An experiment among journalists. Journalism, 21(5), 597-613.

Menkhoff, L., Schmidt, U., & Brozynski, T. (2006). The impact of experience on risk taking, overconfidence, and herding of fund managers: Complementary survey evidence. European Economic Review, 50(7), 1753-1766.

Menton, J. (2020, February 27). Dow plunges 1,191 points, its biggest one-day point drop, as coronavirus fears escalate. Retrieved from https://www.usatoday.com/story/money/2020/02/27/dow-markets-plunge-correction-coronavirus/4889977002/

Metz, C. (2017, June 3). Wall Street Rewired: Apple, Icahn, and the $17 Billion Tweet. Retrieved February 9, 2020, from https://www.wired.com/2013/08/icahn-apple-tweet/

Microsoft Corporation (MSFT) Stock Price, Quote, History & News. (2020, April 21). Retrieved April 21, 2020, from https://finance.yahoo.com/quote/MSFT?p=MSFT

Mihet, R., & NYU, S. E. (2020). Who Benefits from Financial Innovation?.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. The Journal of finance, 32(4), 1151-1168.

Millhiser, I. (2020, March 16). The Dow Jones had its biggest point drop in history Monday. Retrieved March 19, 2020, from https://www.vox.com/2020/3/16/21182341/dow-jones-biggest-point-drop-coronavirus-3000

Mitchell, M. L., & Mulherin, J. H. (1994). The impact of public information on the stock market. The Journal of Finance, 49(3), 923-950.

Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf), 15.

Mueller, G. (n.d.). Why Twitter is the Ideal Platform for Engagement. Retrieved December 26, 2019, from https://www.convinceandconvert.com/social-media-strategy/twitter-engagement/.

Mussa, M. (2000). Factors driving global economic integration. Global economic integration: Opportunities and challenges, 9-55. ISO 690

Nan, A., Perumal, A., & Zaiane, O. R. (2020). Sentiment and Knowledge Based Algorithmic Trading with Deep Reinforcement Learning. arXiv preprint arXiv:2001.09403.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH (1, 1) model. Econometric theory, 6(3), 318-334.

Nguyen, T. H., & Shirai, K. (2015, July). Topic modeling based sentiment analysis on social media for stock market prediction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 1354-1364).

Oh, C., & Sheng, O. (2011, December). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In Icis (pp. 1-19).

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Systems with Applications, 73, 125-144.

Ostrom, C. W., Jr. (1990). Time series analysis: Regression techniques (2nd ed.). Thousand Oaks, CA: Sage Publications.

Pancer, E., & Poole, M. (2016). The popularity and virality of political social media: hashtags, mentions, and links predict likes and retweets of 2016 US presidential nominees' tweets. Social Influence, 11(4), 259-270.

Petram, L. (2014). The world's first stock exchange. Columbia University Press.

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. Journal of Accounting Research, 1-41.

Porshnev, A., Redkin, I., & Shevchenko, A. (2013, December). Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis. In 2013 IEEE 13th International Conference on Data Mining Workshops (pp. 440-444). IEEE.

Porter, M. E., & Kramer, M. R. (2019). Creating shared value. In Managing sustainable business (pp. 323-346). Springer, Dordrecht.

Poterba, J. M., & Summers, L. H. (1984). The persistence of volatility and stock market fluctuations (No. w1462). National Bureau of Economic Research.

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. Journal of Informetrics, 3(2), 143-157.

Rachael , R. (2016, December 6). Donald Trump tweets about Boeing, company instantly loses $1bn. Retrieved February 9, 2020, from https://www.independent.co.uk/news/business/donald-trump-targets-boeing-shares-fall-twitter-air-force-one-planes-dennis-muilenburg-a7458946.html

Rajabi, Z., Uzuner, O., & Shehu, A. Beyond Binary Sentiments: A Multi-Channel BiLSTM-CNN model for Multilabel Emotion Classification of Informal Text.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. PloS one, 10(9), e0138441.

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

Rao, T., & Srivastava, S. (2014). Twitter sentiment analysis: How to hedge your bets in the stock markets. In State of the art applications of social network analysis (pp. 227-247). Springer, Cham.

Reddy, A. (2019, March 12). Boeing's market value has plunged by $40 billion from its 2019 peak (BA) | Markets Insider. Retrieved April 13, 2020, from https://markets.businessinsider.com/news/stocks/boeing-stock-price-market-cap-down-45-billion-from-peak-2019-3-1028024507

Reinganum, M. R. (1981). Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. Journal of financial Economics, 9(1), 19-46.

Revesz, R. (2017, January 6). Donald Trump tweets about Toyota, company instantly loses $1.2bn. Retrieved February 9, 2020, from https://www.independent.co.uk/news/world/americas/toyota-12bn-value-plummet-shares-stock-market-donald-trump-tweet-move-mexico-tax-a7512096.html

Ro, S. (2016, August 17). Past performance is not an indicator of future outcomes for 92.7% of mutual funds. Retrieved March 22, 2020, from https://finance.yahoo.com/news/mutual-fund-past-performance-scorecard-171510113.html

Robehmed, N. (2019, March 7). At 21, Kylie Jenner Becomes The Youngest Self-Made Billionaire Ever. Retrieved February 9, 2020, from https://www.forbes.com/sites/natalierobehmed/2019/03/05/at-21-kylie-jenner-becomes-the-youngest-self-made-billionaire-ever/#4a6eb00c2794

Rojas‐Barahona, L. M. (2016). Deep learning for sentiment analysis. Language and Linguistics Compass, 10(12), 701-719.

Ross, S. A., Westerfield, R., Jordan, B. D., & Biktimirov, E. N. (2004). Essentials of corporate finance. McGraw-Hill/Irwin.

Sadka, R., & Scherbina, A. (2007). Analyst disagreement, mispricing, and liquidity. The Journal of Finance, 62(5), 2367-2403.

Sayrs, L. W. (1989). Pooled time series analysis. Newbury Park, CA: Sage Publications.

Scarf, H. (1967). On the computation of equilibrium prices (No. 232). New Haven, CT: Cowles Foundation for Research in Economics at Yale University.

Shapiro, A. H., Sudhof, M., & Wilson, D. (2020, March). Measuring news sentiment. Federal Reserve Bank of San Francisco.

Shefrin, H. (2002). Beyond greed and fear: Understanding behavioral finance and the psychology of investing. Oxford University Press on Demand.

Shell, A. (2018, August 2). Apple makes history by becoming first US company to reach $1 trillion market value. Retrieved January 10, 2020, from https://www.usatoday.com/story/money/2018/08/02/apple-first-stock-hit-1-trillion-market-value/877867002/.

Smith, A. N., Fischer, E., & Yongjian, C. (2012). How does brand-related user-generated content differ across YouTube, Facebook, and Twitter?. Journal of interactive marketing, 26(2), 102-113.

Social Market Analytics. (n.d.). Retrieved from https://www.socialmarketanalytics.com/

Steger, U. (2003). Corporate diplomacy: The strategy for a volatile, fragmented business environment. John Wiley & Sons.

Stern, G. H., & Feldman, R. J. (2004). Too big to fail: The hazards of bank bailouts. Brookings Institution Press.

Stewart, E. (2019, September 9). The Volfefe Index, Wall Street's new way to measure the effects of Trump tweets, explained. Retrieved February 9, 2020, from https://www.vox.com/policy-and-politics/2019/9/9/20857451/trump-stock-market-tweet-volfefe-jpmorgan-twitter

Stigler, G. J. (1957). Perfect competition, historically contemplated. Journal of political economy, 65(1), 1-17.

Stocks traded, total value (current US$). (2019). Retrieved January 13, 2020, from https://data.worldbank.org/indicator/CM.MKT.TRAD.CD?name_desc=false

Surowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. Economies, Societies and Nations, 296.

Sutcliffe, A. (1994). Time‐series forecasting using fractional differencing. Journal of Forecasting, 13(4), 383-393.

Taylor, M. P., & Allen, H. (1992). The use of technical analysis in the foreign exchange market. Journal of international Money and Finance, 11(3), 304-314.
http://money.visualcapitalist.com/all-of-the-worlds-stock-exchanges-by-size/

The Intercontinental Exchange. (n.d.). Retrieved from https://www.nyse.com/market-cap

Thomson, M. (2006). Human brands: Investigating antecedents to consumers' strong attachments to celebrities. Journal of marketing, 70(3), 104-119.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. Marketing Science, 31(2), 198-215.

Treem, J. W., & Leonardi, P. M. (2013). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. Annals of the International Communication Association, 36(1), 143-189.

Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. Financial Analysts Journal, 57(3), 41-51.

Twitter Usage Statistics. (2020, March 3). Retrieved March 3, 2020, from https://www.internetlivestats.com/twitter-statistics/

Van Deventer, D. R., Imai, K., & Mesler, M. (2013). Advanced financial risk management: tools and techniques for integrated credit risk and interest rate risk management. John Wiley & Sons.

Wang, X., Yu, C., & Wei, Y. (2012). Social media peer communication and impacts on purchase intentions: A consumer socialization framework. Journal of interactive marketing, 26(4), 198-208.

Weber, M., Weber, E. U., & Nosić, A. (2013). Who takes risks when and why: Determinants of changes in investor risk taking. Review of Finance, 17(3), 847-883.

Wei, X. (2016) Media Bias, Investor Psychological Bias, and the Cross-Section of Stock Returns.

What Price Yield? Fund Managers 'Taking Excessive Risk'. (2015, April 30). Retrieved from
https://www.ai-cio.com/news/what-price-yield-fund-managers-taking-excessive-risk

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: journal of the Econometric Society, 817-838.

Williams, C. (2007). Research methods. Journal of Business & Economics Research (JBER), 5(3).

Williams, E. E., & Dobelman, J. A. (2020). Beyond the CAPM. World Scientific Book Chapters, 137-160.

Wong, F. M. F., Tan, C. W., Sen, S., & Chiang, M. (2013, June). Quantifying political leaning from tweets and retweets. In Seventh International AAAI Conference on Weblogs and Social Media.

Woo, K. Y., Mai, C., McAleer, M., & Wong, W. K. (2020). Review on efficiency and anomalies in stock markets. Economies, 8(1), 20.

Woods, H. C., & Scott, H. (2016). # Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. Journal of adolescence, 51, 41-49.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. Decision Support Systems, 55(4), 919-926.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through twitter "I hope it is not as bad as I fear". Procedia-Social and Behavioral Sciences, 26, 55-62.

Zimbardo, P. G., & Leippe, M. R. (1991). The psychology of attitude change and social influence. Mcgraw-Hill Book Company

Zhou, Z., Gao, M., Liu, Q., & Xiao, H. (2020). Forecasting stock price movements with multiple data sources: Evidence from stock market in China. Physica A: Statistical Mechanics and its Applications, 542, 123389.

Appendices

Data Example:

Date	Username	Retweets	Text	Sentiment	Binary	Favorites
12/30/2018	BillGates	1377	I hope you	p	1	13062
12/30/2018	BillGates	243	In the mea	o	0	4572
12/30/2018	BillGates	331	Although I	o	0	1858
12/30/2018	BillGates	454	Gene editi	o	0	1603
12/30/2018	BillGates	139	The good	p	1	884
12/30/2018	BillGates	135	I had hope	n	-1	759
12/30/2018	BillGates	258	Some peo	n	-1	1022
12/30/2018	BillGates	93	This year (	o	0	657
12/30/2018	BillGates	116	Almost a c	p	1	795
12/30/2018	BillGates	128	Just over a	p	1	1021
12/30/2018	BillGates	2643	1/ What a	o	0	14138
12/22/2018	BillGates	2120	The books	o	0	10753
12/21/2018	BillGates	1483	I'm excite	p	1	9826
12/20/2018	BillGates	486	New repo	o	0	3152
12/20/2018	BillGates	1031	This year I	p	1	9384
12/20/2018	BillGates	343	New data.	p	1	2559

Date	Positive	Neutral	Negative
1/4/2016	0	0	0
1/5/2016	0	0	0
1/6/2016	0	1	0
1/7/2016	1	0	0
1/8/2016	1	0	0
1/11/2016	0	1	0
1/12/2016	1	0	0
1/13/2016	0	0	0
1/14/2016	0	2	0
1/15/2016	0	1	0
1/19/2016	0	0	0
1/20/2016	1	0	0
1/21/2016	0	1	0
1/22/2016	0	1	0
1/25/2016	0	2	0
1/26/2016	0	1	0

Organization Example:

Date	S&P 500	Users	Retweets	Sentiment		Positive	Neutral	Negative
1/4/2016	2012.660034	9	477078	0.667		15	22	0
1/5/2016	2016.709961	10	146274	0.8		21	38	2
1/6/2016	1990.26001	9	356938	0.778		18	36	4
1/7/2016	1943.089966	9	347352	0.444		16	31	7
1/8/2016	1922.030029	8	348345	0.875		24	33	4
1/11/2016	1923.670044	12	309568	0.5		23	30	6
1/12/2016	1938.680054	10	315857	0.6		32	51	4
1/13/2016	1890.280029	9	240570	0.556		40	43	14
1/14/2016	1921.839966	10	228451	0.5		23	37	5
1/15/2016	1880.329956	8	269417	0.25		9	32	3
1/19/2016	1881.329956	10	113063	0.4		29	22	7
1/20/2016	1859.329956	7	191799	0.714		18	36	2
1/21/2016	1868.98999	9	262046	0.556		19	29	6
1/22/2016	1906.900024	9	217509	0.333		12	38	13
1/25/2016	1877.079956	10	413698	0.4		10	39	7