2020

Customer and Employee Social Media Comments/Feedback and Stock Purchasing Decisions Enhanced by Sentiment Analysis

Drew Mikel Hall
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral study by

Drew Mikel Hall

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Jaime Klein, Committee Chairperson, Doctor of Business Administration Faculty

Dr. Gwendolyn Dooley, Committee Member, Doctor of Business Administration Faculty

Dr. Irene Williams, University Reviewer, Doctor of Business Administration Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2020

Abstract

Customer and Employee Social Media Comments/Feedback and Stock Purchasing

Decisions Enhanced by Sentiment Analysis

by

Drew Mikel Hall

MBA, University of Maryland Global Campus, 2017

BS, University of Maryland Global Campus, 2015

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Business Administration

Walden University

December 2020

Abstract

The U.S. Securities and Exchange Commission (SEC) warns professional investors that sentiment analysis tools may lead to impulsive investment decision-making. This warning comes despite evidence showing that aided social sentiment investment decision tools can increase accurate investment decision-making by 18%. Using Fama's theory of efficient market hypothesis, the purpose of this quantitative correlational study was to examine whether customer Twitter comments and employee Glassdoor feedback sentiment predicted successful investing decisions measured by business stock prices. Two thousand records from 3 archival U.S. public NASDAQ 100 datasets from March 28, 2016, to June 15, 2016 (79 days) of 53 companies with over 100 comments were analyzed using multiple linear regression. The multiple regression analysis results indicated no significant predictability for successful investing decisions, $F(10, 2993) = .295$, $p = .982$, $R^2 = .001$. The results indicated that the sentiment from both Twitter and Glassdoor was not necessarily an indicator for investors to make successful investment decisions for the 79 days in 2016. The knowledge about Artificial Intelligence (AI) sentiment usage may help professional investors gain profit or prevent losses. A recommendation to investors is to heed warnings from the SEC about tools for sentiment analysis investment decisions. Implications for positive social change include preventing an investor from using a risky sentiment tool for investment decision-making that may lead to losing capital.

Customer and Employee Social Media Comments/Feedback and Stock Purchasing

Decisions Enhanced by Sentiment Analysis

by

Drew Mikel Hall

MBA, University of Maryland Global Campus, 2017

BS, University of Maryland Global Campus, 2015

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Business Administration

Walden University

December 2020

Dedication

This study is dedicated to my wife Bethany for her support and willingness to allow me the opportunity to be the first doctor of my blood-related family. I would also like to thank my father and mother for instilling my work ethic, and stepfather for showing me that becoming a doctor is achievable. To my stepmother, thank you for your guidance and unmatched wisdom that has guided me to this achievement. My family is quite amazing, thank you, and I love you all.

Acknowledgments

I would like to thank Dr. Jaime Klein, my chair, Dr. Gwen Dooley, my second committee member, Dr. Irene Williams, my URR, and Dr. George Bradley. Thank you all very much for your guidance, patience, and encouraging words.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study

The efficient market hypothesis (EMH) is not a new theory for economic or financial scholars. Ren, Wu, and Liu (2019) suggested that the EMH has flaws when measuring information sentiment and how it may help investors gain significant profits. However, many articles testing EMH validity have suggested that there is truth to the theory for instances that depend on the variables a researcher may use for analysis (Leekha, Wadhwa, Jain, & Wadhwa, 2018). The purpose of this study, therefore, was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by stock prices.

Background of the Problem

The stock market is a fickle instrument for exchanging property for money. The stock market consists of algorithms and people that inhibit the flow of trends based on new information. Investors use new information in various forms, from social media posts to news articles and insider trading in secret places. The effect of this information, whether it be a positive or negative sentiment, is a broad research subject for academics and financial practitioners alike (Ren et al., 2019). The creation of the world wide web brought about a tremendous amount of information flow known today as *big data*. A large amount of information growth brought new and more significant social platforms every day (Pourkhani, Abdipour, Baher, & Moslehpour, 2019). The information from these social platforms could hold much value for investors to find and use these data for profit. There are current investment tools that use sentiment analysis on large amounts of text from social media platforms to increase accurate investment decision-making by

18% (Cwynar, Cwynar, & Pater, 2017). However, only 31% of professional financial investors use sentiment analysis as an investment decision-making tool (Daniel, Neves, & Horta, 2017). The stock market is an evolving entity, and with it, the analytics necessary to help with safer investment decision-making.

The social media platform that has the most attention from sentiment analysis researchers is Twitter. Twitter posts can come from random consumers, news outlets, influencers, and professional financial advisors (Ramon Saura, Reyes-Menendez, & Alvarez-Alonso, 2018; Xun & Gou, 2017). The accuracy of sentiment from the different types of post-relationship to business stock prices varies widely. Zaeem and Barber's (2019) study sentiment analyzed Twitter posts and company stock price fluctuations with an 80% accuracy. However, Nisar and Yeung (2017) found no significant correlation between Twitter posts' sentiment and forecasting different company stock prices.

Other social platforms may offer safer information for investment decision-making, such as Glassdoor. Chamberlain (2015) is an economist who produced a study with findings that Glassdoor's 100 best companies to work for portfolio outperformed the S&P 500 for 5 straight years with an average of 115% higher returns. However, Sheng (2019) found that using Glassdoor data increased investors' decision-making by 7% to 9% on average. However, sentiment analysis tools may be too risky for potential enhancement of investment decision-making (Pourkhani et al., 2019; U.S. Securities and Exchange Commission, 2019). Investors may be losing out on using sentiment analysis, or they are preventing potential losses.

Problem Statement

The U.S. Securities and Exchange Commission (2019) warns investors that sentiment analysis tools may lead to risky, emotional, or impulsive investment decision-making. This warning comes despite evidence showing that aided investment decision-tools such as those used to measure social sentiment can have an accuracy ranging from 70% to 90% for investment decision-making (Ren et al., 2019). The general business problem was that investors are unsure whether the use of sentiment analysis leads to better investment decisions. The specific business problem was that some investors may not know the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices.

Purpose Statement

The purpose of this quantitative correlational study was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices. The independent variables were customer comments from Twitter on company profiles, Glassdoor Likert scale ratings, and employee comments on company profiles. The dependent variable was successful investing decisions measured by business stock prices. The target population was archival data from U.S. companies listed on the NASDAQ index. The implications for social change might include more effective investment decision-making, which may increase funds for retirements, children's education, or other socially desirable uses.

Nature of the Study

I chose a quantitative methodology for this study. A quantitative method is useful for understanding patterns of connections mathematically between dependent and independent variables (Bernard, 2013). This quantitative study's foundational data was from public sources used to predict the relationships between the independent and dependent variables. A quantitative methodology was appropriate for this study because parametric measurements are useful for identifying generalized relationships (Allwood, 2012). Researchers using a qualitative methodology usually use inductive analysis and interviews to uncover experiences and meaning (Castellan, 2010). This study was not about uncovering experiences or meaning from the variables. Therefore, as this study aimed to discover relationships between variables, the qualitative or mixed-method approaches were not appropriate.

Specifically, the correlation design via multiple regression was appropriate for this study. Researchers often use a correlational design for testing relationships without assuming cause and effect (Bernard, 2013). A correlation research design focuses on independent predictor variables and a target-dependent variable, all measured for a positive or negative trend to determine inference (Tabachnick & Fidell, 2007). If a study has only two independent variables and measuring the difference of the means is necessary, then a *t* test could be employed (Schneider, 2013). However, this study has two independent variables and a single dependent variable, making a *t* test unnecessary. Researchers using an experimental design mostly use ANOVA, ANCOVA, or MANOVA techniques, and assess for causality; in contrast, this study calls for assessing correlation (Bernard, 2013). The purpose of this study was to measure multiple independent variables with one dependent variable. Thus, a correlation design used to measure the relationships between independent variables and a single dependent variable was appropriate for this study.

Research Question

The research question for this study was: What is the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices?

Hypotheses

H_0: There is no statistically significant relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices.

H_1: There is a statistically significant relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices.

Theoretical or Conceptual Framework

I chose the EMH as the theoretical framework for this study. Fama, who conceptualized this theory in 1965, stipulated that relevant information inherently applies to stock prices, making them at fair value at all times (Fama, 1965). The EMH's development from its debut produced naysayers and hypothesis testing with different statistical methods to see if all information is genuinely apparent in the current stock price (Leekha et al., 2018). Fama also proposed that new public information is immediately calculable in the stock price, and beating the market is more about luck than skill (Leekha et al., 2018). However, if one investor undervalues a stock while another investor looks at the potential of growth from the same stock, the fair market value is determined differently (Ding, Mazouz, & Wang, 2019). One investor may be interpreting or not using specific information to account for a stock's market value, which brings the validity of the EMH into question (Ding et al., 2019). With a larger quantity and a different form of measuring information and an ability to influence decision sentiment from social media could impact investor decisions (Cakra & Distiawan-Trisedya, 2015). Social media produces a higher quantity of information than was possible in the past. Depending on the investors' perception of sentiment, it could drive the stock's market value or price (Pourkhani et al., 2019). If the EMH theory were holding, then the use of sentiment analysis would not produce better-investing decisions. However, if sentiment analysis helped predict future stock prices, it would indicate the EMH might not apply in the cases studied.

Operational Definitions

NASDAQ: The National Association of Securities Dealers Automated Quotations (NASDAQ) is a trading platform of over 3,100 securities that investors use to buy and sell electronically (Li, Chan, Ou, & Ruifeng, 2017).

Investment decision-making: Investment decision-making is the choice to buy, sell, or hold assets (Cao, Ouyang, Li, Li, & Chen, 2018).

Artificial intelligence (AI): AI is computerized algorithmic platforms that try to mimic and enhance human intelligence from decision-making to visual perception (Ibrahim & Wang, 2019).

Assumptions, Limitations, and Delimitations

Assumptions

According to Sale, Lohfeld, and Brazil (2015), assumptions are prevalent and necessary for studies involving any trace of data gathering. An assumption in a study thought to be true even without evidence to prove the assertion is apparent (Castellan, 2010). The primary assumption of this study was that the sentiment data gathered was from humans and not bots. Another assumption was that the data obtained would be large enough to provide some inference and validity to the outcome of the study. Finally, I assumed that the Glassdoor reviews are by people that worked at the company, and Twitter users understood what company they referred to when providing public sentiment.

Limitations

The limitations in studies are beyond a researcher's control (Polit & Beck, 2010). Limitations may affect the study creating a different outcome or misinterpretations of the findings. The primary limitation of this study was the use of secondary data that may not be accurate. Another limitation was from using quantitative research that may be too generalized and misses the potential depth of understanding when using a qualitative method. The quantitative study can only show correlation and not causation, which was a limitation. The timeframe and amount of data was another limitation.

Delimitations

The researcher chooses delimitations as a form of boundaries to narrow the study focus (O'Leary, 2017). A delimitation of this study was the focus on only large capital NASDAQ companies. The readers should account for the delimitation when interpreting the findings. However, stock prices and sentiment are elements used in many different country stock market indexes.

Significance of the Study

Value to the Business

The value of this study to investors is that it may help increase knowledge of sentiment analysis and investing decisions from customers and employees' posts and feedback. Many previous studies used sentiment analysis, looking at the correlation between positive or negative sentiment expressed on different social media platforms and stock prices. Still, the results have no direct conclusive correlation between sentiment and fluctuations in stock price (Pourkhani et al., 2019). The value to the investor's knowledge about AI sentiment usage may help with gaining profit or preventing losses.

Contribution to Business Practice

Identifying the importance of sentiment analysis for investor decision-making could increase optimism and risk-taking. Paraboni, Righi, Vieira, and Silveira (2018) identified that significant markets from the United States and Germany view sentiment as a drive for optimism and risk barring investment decision-making. Yadav, Kumar, and Kumar (2019b) suggested that investors who can decipher sentiment faster than other investors in the active market when making investment decisions can make a more substantial profit than others. Investors wary of using sentiment analysis AI as a business practice could prevent losses or gain more profits by having a relationship identifiable between them.

Implications for Social Change

This study's potential positive social change may improve investment decision-making, leading to profit, philanthropy, and increasing the cash flow to a local economy. Increasing knowledge of sentiment analysis effectiveness may improve investors' success rate and increase shareholder value. The change could lead to the preservation of retirement savings from IRAs and 401ks that would benefit many people.

A Review of the Professional and Academic Literature

The purpose of this quantitative correlational study was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions. The independent variables were customer comments from Twitter on company profiles, Glassdoor Likert scale ratings, and employee comments on company profiles. The dependent variable was the outcome of investing decisions measured by stock prices. I reviewed the EMH, which was the theoretical framework of my study. The theory has multiple lenses from a weak to a strong form, depending on the different types of information and its potential effect on stock price variances (Jovanovic, 2018). The EMH maintains that a stock price is always currently allocated correctly at any point in time. No new information will give any investor the ability to gain a significant profit (Leekha et al., 2018). A market's efficiency as being emerging or mature plays a part in Fama's EMH theory validity.

U.S. Stock Market

The U.S. market is a capital trading platform for new and existing stocks to be bought and sold (Machlup, 1940; Worthington, 2013). The buying and selling entities are called stock exchanges or stock markets, where those who have ownership of stocks and bonds and those who wish to buy stocks can exchange currency for ownership (Worthington, 2013). The stock market's primary purpose is to allow firms to gain financing from their stock offering. The firms can eventually gain more capital when selling additional stocks or splitting the stocks to create a larger quantity of available stocks to trade (Machlup, 1940). The stock market serves as a resource for businesses by providing investment to the firms and what to expect in terms of capital cost when buying, selling, or splitting stocks (Machlup, 1940). Furthermore, stock prices will depend on the investors' forecast of a company's performance, with the added effect of supply and demand driving the trading price.

The foresight of investor decision-making is driven by emotional bias at times, which is subjectively a motivator to expect a company to do well or not (Chinas, 2018). The stock will then either gain or fall in price, leaving the investor with a potential profit or loss (Chinas, 2018). The loss and profit potential are emotional drivers, and AI tools may help or hinder the investment decision process for investors.

This literature review consists of six sections: Purpose statement; hypotheses; EMH theory; Twitter comments; and Glassdoor Feedback as independent variables, investment decision-making measured by stock prices as a dependent variable, and the measurement of variables. The first section addresses the purpose of this ex post facto correlational study, followed by a restatement of the hypotheses. The next section provides an in-depth analysis and synthesis of EMH theory literature, including other supporting and contrasting theories. The following section addresses the relevance and contrasting nature of EMH theory through critical analysis and synthesis of the independent and dependent variables. A review of the independent and dependent variables' measurement follows an examination of the independent and dependent variables. The final section compares and contrasts differing points of view and their relationship to previous research and findings.

Literature Review Organization and Strategy

The literature review's organization and strategy may help provide readers with an understanding of the structure of the review content. The next section includes an in-depth review of the theoretical framework, followed by a review comparing and contrasting rival theories and measurement methods. The literature review concludes with a review and explanation of the predictor variables I used in the study.

I used multiple electronic databases for online peer-reviewed journal articles and other materials on the study's nature. I searched using ProQuest, EBSCO, IEEE, Walden University Library, Cornell's Arvix.com, and Google Scholar. I searched using Boolean parameters and peer-review setting for articles that included the terms: *sentiment analysis*, *EMH*, *SVM*, *Naïve Bayes*, *stock market*, *stocks*, *stock price*, *neural networks*, *convolutional neural networks*, *Monkeylearn*, *MITSMR Culture 500*, *sentiment algorithms*, *Twitter*, *Glassdoor.com*, *political news*, *news*, and *investment decision-making*. The literature search output ranged from eight to 980. I found 112 relevant articles published between 2015-2020, and 107 of those articles are peer reviewed. Table 1 provides a depiction of the content from the citations found in the literature review.

Table 1

Literature Review Citations

Type	Citations	Peer-reviewed	% Peer review	Published within 5 years	Percent within 5 years
Journal article	110	107	97%	94	85%
Book	1	0	0%	1	100%
Dissertation	2	0	0%	2	100%
Total	113	107	97%	97	95%

Purpose Statement

The purpose of this quantitative correlational study was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices. The independent variables were customer comments from Twitter on company profiles, Glassdoor Likert scale ratings, and employee comments on company profiles. The dependent variable was successful investing decisions measured by business stock prices. The target population was archival data from U.S. companies listed on the NASDAQ index. The implications for social change might include more effective investment decision-making, which may increase funds for retirements, children's education, or other socially desirable uses.

Hypotheses

H_0: There is no statistically significant relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices.

H_1: There is a statistically significant relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices.

Efficient Market Hypothesis

I chose the EMH as the theoretical framework for this study. Fama conceptualized the theory in his 1965 PhD dissertation, stipulating that relevant information inherently applies to stock prices, making them fair at all times in a perfectly efficient market. The development of the theory came from economists trying to explain the randomness of stock market prices (Jovanovic, 2018). The ideas of stock price movement having a predictable pattern and how information influences prices were first discussed in 1863 by a Frenchmen named Regnault and a mathematician in the 1900s who created a mathematical model to test this idea (Jovanovic, 2018). The model was called Brownian motion, which led to the formulation of two other models called the random walk and martingale in the 1960s. These models began the race to theorize stock price changes and how information may affect such price movements. Fama (1970) used the martingale model to test a hypothesis that all available information used would produce a zero-sum profit. Fama's theory has been widely tested for accuracy and debate. The EMH theory includes new public information, as well. If someone makes an abnormal return from public information, it is luck rather than skill (Leekha et al., 2018). However, many economists have challenged the EMH and Fama's definition and choice of mathematical models.

There are extensive debates pertaining to the validity of EMH. Economists Leroy, Lucas, and Samuelson disputed EMH in the 1970s by analyzing the theoretical proof showing stock price tendencies to be equal to the randomly fluctuating intrinsic values of individuals (Jovanovic, 2018). Grossman and Stiglitz in the 1980s postulated that informationally efficient markets are impossible because all available information cannot correctly reflect prices (Gillet & Renault, 2019). The perceivable scope of EMH began to be more debatable, depending on a given economist's data and analysis method, including the notion that only specific sets of information testable with EMH provides that a set of information may prove to be true or false. Not all public information is accountable at any one time to provide inferential validity.

As information comes in droves in the 21st century, the past economist assertions may have less validity today. Malkiel and Fama's (1970) EMH study indicated that securities prices fully reflect all publicly available information. However, in 2003, Shiller found that no study at the time ever supported the EMH linking stock market fluctuations for the aggregate stock market. The validity and applicability of the theory became a debatable subject for many years. In 2003, Malkiel and other economists challenged the EMH to explain diverging market sentiment, including how psychological and behavioral elements impact stock prices. The evolution of the EMH theory definition and applicability became a common test ground for economic researchers.

The EMH has three differential tiers for the different types of perceivable market environments: weak, semi-strong, and strong. The weak form of efficiency is the notion that the past prices of any stock will never predict future prices (Chinas, 2018; Hawaldar, Rohit, & Pinto, 2017; Nwachukqu & Shitta, 2015). Semi-strong efficiency refers to any form of public information already allocated in the stock price (Chinas, 2018; de Aragão Batista, Maia, & Romero, 2018; Nwachukqu & Shitta, 2015). No one can use such public information to make an egregious profit (Chinas, 2018; de Aragão Batista et al., 2018; Nwachukqu & Shitta, 2015). Strong form efficiency is when all public and insider trading information is accountable in the current stock prices, which means there is no way of gaining a substantial profit because it already reflects all available information (Alekneviciene, Kviedaraitiene, & Alekneviciute, 2018; Chinas, 2018; Nwachukqu & Shitta, 2015). The three different EMH elements cover the majority of potential markets from the formulation of a market to its most mature point in time.

Weak form. The weak form of efficiency is that no past prices of any stock will ever predict the future price. Hawaldar et al. (2017) tested the weak form of EMH with a Bahrain stock portfolio of 43 companies and found that seven of those companies' stock prices were predictable based on their past prices for up to 4 years. Moreover, Kršikapa-Rašajski and Rankov-Siniša's (2016) study on the Baltic stock exchange, using 1 year of stock price data, provided similar predictability with past stock prices that also rejects the weak EMH theory. However, there is a suggestion that most emerging markets may start with no efficiency. With time and successful growth, however, the market will attain a present weak form of EMH. (Fama, 1970). Andrianto and Mirza's (2016) correlation study of the Indonesian major stock exchanges with 1 year of stock price data from 2013-2014 and a random walk test for the weak form of EMH found a strong relationship with the market having a weak form of efficiency. Andrianto and Mirza suggested that the market slowly grew more efficient by maturing market practices and a greater amount of capital investment. Nwachukqu and Shitta (2015) looked at 24 emerging markets stock prices for 10 years from 2000-2010 from all regions of the world with a random walk analysis to test the weak form of EMH. Nwachukqu and Shitta found that 18 of the markets do not have a weak form of efficiency.

Furthermore, these studies were on markets believed to be newer and more volatile compared to the mature markets of the United States, Japan, Germany, and the United Kingdom, among others. Nwachukqu and Shitta also measured nine mature economies from 2000-2010 with a random walk analysis to the weak form of EMH. The study findings suggested that all nine mature industries have a weak form efficiency, with at least three of them exhibiting short periods of stock price predictability for a random number of days that collectively equates to six months out of the ten years of data. The evidence that a weak form of efficiency can exist in mature and emerging markets brings the validity of semi-strong market efficacy into question.

Semi-strong. A semi-strong efficiency is that all public information will be reflected in the real-time stock price, meaning gaining a significant profit is not possible when using public information. Alekneviciene et al.'s (2018) study on the Baltic exchange using over 16 years of data to test the semi-strong EMH model via measuring abnormal returns found that the exchange was inefficient. Alekneviciene et al. (2018)

suggested that many investors could make a significant return from the stocks on the Baltic exchange. de Aragão Batista et al.'s (2018) study of the Brazilian stock exchange and the semi-strong EMH theory for 2016 during a presidential impeachment found a correlation with a semi-strong efficient stock market. The efficiency of a stock market may have different correlations depending on many factors of information and the market's maturity. Gyamfi, Kyei, and Gill's (2017) study of the African stock market semi-strong EMH correlation studies from 1974-2016 findings suggested no strong correlation with semi-strong efficiency in any of the markets. Gymafi et al. also stated that potential publication bias may be present in the body of research. With the symptom of potential study bias, present validation, and reliability lessen.

In contrast, Manasseh, Ozuzu, and Ogbuabor's (2016) study of the Nigerian stock market from 2002-2006 via event study methodology findings on 121 information releases of events had a relationship to supporting strong efficiency only in 2002. Moreover, the small capital stocks or penny stocks were affected the most. In contrast, medium-sized and large capital stocks were not affected (Manasseh et al., 2016). The different market capitalizations of companies' stocks may be a dividing factor to EMH validity. However, Ayodele, Anthony, and Adekemi's (2016) study of the Nigerian stock market with randomly sampled data from 2005-2013 and all small-, medium-, and large-capitalization stocks in the population found that public information has a relationship to all sizes of stock capitalization prices and suggested the market has semi-strong efficiency. The size of stock capitalization and even the overall market's maturity may have a relationship to EMH validity.

Strong form. The strong form of EMH is testing private information with all other available information. de Aragão Batista et al.'s (2018) study of the Brazilian stock exchange also tested for the strong form of EMH with findings rejecting the hypothesis because the information known by insider traders was the same as the public information. Rani, Kaushal, and Shakir (2019) suggested that Twitter influences business and corporate information leakage among the Fortune 500 companies because 77% of the employees are actively tweeting throughout the day.

Evaluating internal information and testing the strong form of EMH is quite difficult for many researchers to publish findings because of insider trading implications (Lekovic, 2018). However, Kumari (2019) found a way to measure insider trading knowledge through bonus announcements from 2014-2019 in 9 large banking companies in India. The study findings suggested that even with bonus information staying internal and later being accessible to the public, it has no relationship to stock price variances, suggesting that the nine Indian businesses' stock prices may have strong efficiency. Private and public information may not have a different relationship when being used for investment decision-making.

The EMH is the framework for several correlation studies relating to measuring investment decisions making by information usage and stock prices. The following section delves into the relationships between EMH and stock price predictability measured by sentiment analysis. The review of the EMH framework concludes with a discussion of the relationship to investment decision-making.

Sentiment Theory

The sentiment itself refers to the negative, neutral, or positive belief in context, either heard or written. According to sentiment theory, many subjective written instances, either publicly available or privately shared, may affect decision-making (Gao, Ren, & Zhang, 2020). The theory specific to economics concerns the presence of information having a negative, neutral, or positive context affecting stock prices. If there is no information present, the stock prices will revert to the original price (Gao et al., 2020; Zhou, Xu, & Zhao, 2018). The most recent literature uses sentiment theory to measure sentiment with a popular social media app called Twitter and measuring the relationship to stock prices and customer decision-making around the world.

Twitter is a social application for posting up to 280-character messages to the 250 million+ users of the platform (Ibrahim, Wang, Bourne, 2017; Sheng, Amankwah-Amoah, & Wang, 2017). The use of sentiment classification on social media posts has no perfect generalized criteria for evaluating subjective sentiment regarding positive, neutral, or negative expressions (Ibrahim & Wang, 2019). However, the first computerized sentiment analysis tool debuted in the early 2000s and began the innovation to generalize the practice of accurately analyzing sentiment.

Sentiment Analysis

Humans have used sentiment analysis when reading or listening to conversations by mapping words, tones, and contexts to understand sentiment output. Sentiment analysis methods, techniques, and tools can detect and extract subjective information such as opinions and attitudes from language (Mäntylä, Graziotin, & Kuutila, 2018). One of the earliest studies of sentiment analysis came from a pre-World War II public opinion collection on socialism (Mäntylä et al., 2018). However, today the innovations of artificial intelligence (AI) have brought about algorithm-driven sentiment tools that involve software development languages to develop and build accurate predictions. An AI sentiment tool can recognize what is known as unstructured text meaning from typed human sentences and analyze this type of data on a giant scale (Daniel et al., 2017; Kim, Jeong, Kim, Kang, & Song, 2016). The AI is trainable when matching analyzed unstructured text to a list of predefined words determined as being positive or negative (Ibrahim & Wang, 2019). Once the computer evaluates the information with a predefined library of positive and negative words, a human is still necessary to train the computer to distinguish potential sarcasm from negative words that could mean something positive (Gao et al., 2020; Zhou et al., 2018). The most recent use of sentiment analysis AI is testing on 280 characters unstructured text from many Twitter posts.

The most recent advancement for sentiment analysis algorithms is significantly impacting accuracy for depicting human sentiment. A type of sentiment analysis lexicon technique helps to better evaluate the Twitter post sentiment from many areas such as retail, finance, and airlines, producing a higher accuracy of sentiment scoring output (Daniel et al., 2017; Guo, Vargo, Pan, Ding, & Ishwar, 2016; Kim et al., 2016; A. Yildirim, Sküdarlı, & Zgür, 2016). Kim et al.'s (2016) study showed that Twitter users express their interests and attitudes about news updates, and 19% of the tweets contain brand sentiment, both positive and negative. Researchers even used sentiment analysis to derive relationships between financial market predictions and customer decision-making.

Bollen, Mao, and Zeng's (2011) study measuring sentiment from randomly sampled Dow Jones Industrial Average company Twitter posts and time-series stock price variances findings suggested they have a significant relationship (Bollen et al., 2011). A study by Liau and Tan (2014) of 10,895 Tweets about airlines in Malaysia revealed four main topics: ticket promotions, customer service, flight cancellations, and pre-booking management as the most sentiment-derivable posts for decision-making. Liau and Tan's study produced business intelligence on the relationship of these topics to customer decision-making. Chu and Sung (2015) found a correlation with sentiment influencing retweeting behaviors. Researchers have even used sentiment analysis output on different publicly traded companies' profiles to discern a potential relationship between stock prices and index valuations.

Investors that use professional financial advice can come from television to social media posts. Daniel et al.'s (2017) study on financial tweet sentiments from experienced personnel in the financial industry correlated to successful financial decisions for investors such as profit gains and preventing losses. Daniel et al. also found that 80% of financial professionals use social media, with only 31% using social media sentiment information for investment decisions. Even with forecasting projections showing an 18.6% accuracy increase for successful decision-making when sentiment is incorporated (Cwynar et al., 2017). The professional financial industry may have a cautionary view of sentiment analysis tools and its potential ability to help with investment decision-making.

The information and sentiment on different companies can have a relationship with its stock price fluctuations. Public information that directly impacts a company may influence people's decisions to buy and sell stocks (Heston & Sinha, 2017; Nisar & Yeung, 2017; Yoosin, Rahul, Jie, Seung, 2013). Moreover, people tend to buy company stocks with a good reputation (Yoosin et al., 2013). One way to know a company's reputation is by seeing the relationship of sentiment between the company and its customers. One way to derive such information may be from social media posts that pertain to the moods that affect customer decision-making.

With enough social media posts projecting a specific sentiment, the probability of a relationship to stock market performance increases. Bollen et al. (2011) found that certain moods from Twitter posts could predict the Dow Jones Industrial Average. However, Nisar and Yeung (2017) found no significant correlation between the sentiment from Twitter posts and forecasting different company stock prices. The findings of different studies suggest that sentiment may or may not have a relationship to stock price fluctuations. Moreover, whether negative, positive, or neutral, the quantity and type of information may correlate with specific company stock prices (Leekha et al., 2018). Whether information immediately created by customers on Twitter reflects the current price has wavering correlations, and no one definitive answer (Nisar & Yeung, 2017). If there is a relationship from information, it seems to have no or a delayed instance of predicting stock prices (Heston & Sinha, 2017). The use of sentiment information from social media platforms is still in the wavering phases of validation and lack of reliability.

Review of Alternative Theories

Noise trader theory. A quick and emotionally driven decision maker is what some consider to be a noise trader. The opposing theory to EMH is the noise trader theory because of quick asset price changes based on investor behaviors affected by emotional and knee-jerk reaction trades from new investors (Ding et al., 2019). Moreover, the noise trader theory is measurable with a positive and negative volume index that is useful to traders (Ding et al., 2019). These indexes are used by professional traders as a compass to verify if noise traders are manipulating prices via the volume of stock movement over a day or longer (Ding et al., 2019). Thus, this theory begins with a similar approach of using sentiment to aid decision-making.

Noise trading and stock price. Noise traders and stock price fluctuations are a phenomenon that professional investors believe to be true. Bergsma and Tayal's (2019) study measures noise trader theory with stock price changes and the potential to gain large profits. Bergsma and Tayal stated that noise trading has a short effect on the overvaluing and undervaluing of individual risky stocks. Sinha's (2015) study of the India NIFTY 50 index and the effects of noise trading from one to five days showed a positive relationship between noise trading and stock price fluctuations. However, once a specific time has passed, the stock prices begin to show lower volatility (Bouteska, 2019). The noise trader theory is another lens for viewing investment decision-making based on stock price variances. However, the theory is not concerned with the sentiment of the information, but rather on the specific body of investors driven by emotional responses to buy and sell stocks.

Prospect theory. The potential negative information of companies may have a relationship with investors' decision-making. Kahneman and Tversky's (1979) economic prospect theory suggested that negative news has a prolonged and more significant effect on investment decisions than positive news (Chen, Luo, Liu, & Zhang, 2018). Chen et al. (2018) used prospect theory as a theoretical lens and sentiment analysis to evaluate the impacts of negative and positive information on investors' stock purchasing decisions. Paraboni et al. (2018) completed a study using sentiment analysis of company information, stock prices, and risk aversion for investor decision-making with a positive relationship founded in prospect theory. Cao et al., (2018) used prospect theory as a lens to evaluate stock price jumps by simulating different traders such as noise, day, and long-term traders and their investment decision-making. Cao et al.'s simulation findings suggested a relationship between price jumps from the different types of traders and information they all use to make decisions. Rahayu (2017) focused on 192 students' investment decision-making when given positive and negative dividend information in different market environments. The study's findings suggested that in a bull market, both negative and positive information have a greater relationship to changes in investment decisions. The prospect theory is not an applicable theory to this study because the research question does not pertain to the length of time for negative or positive effects from information on investing decisions.

Adaptive market hypothesis. A newer proposed theory called adaptive market hypothesis (AMH) is the leading theory that expands on the EMH perimeters. Lo (2004) developed an alternative theory to the traditional EMH (Lekhal & Ahmed, 2020; Shahid, Jehanzeb, Abbas, Zubair, & Akbar, 2019). The AMH framework reconciles market efficiency with principles of evolution, adaptation, competition, and natural selection in terms of investment decision-making (Lekhal & Ahmed, 2020; Shahid et al., 2019). Lo suggested that investors are rational and choose good or bad investments from market conditions such as market evolution, financial technologies, and survival of the fittest practices (Lekhal & Ahmed, 2020; Shahid et al., 2019). When market conditions change and information technology advances can cause fluctuations in prices, this is an indicator of the AMH theory in practice (Lekhal & Ahmed, 2020). AMH theory is beyond the scope of this study because there is a lack of mathematical models available to justify relationships or causation.

Measurement of Variables

Naïve Bayes

The Bayesian Theorem is a decision algorithm used in many different studies from statistical theories in biological studies to evaluate sentiment from text. The unique characteristics of this theorem are that it analyzes each word in a sentence as an individual variable and uses predefined values for each word to produce an average mean for the sentences as a whole (Song, Kim, Lee, Kim, & Youn, 2017; Jadon, Bhatia, & Mishra, 2019). The Bayesian approach has some weaknesses because of stop words such as and, it, the, and is that affect the accuracy of sentiment analysis output. Studies that use only Naïve Bayes for sentiment analysis widely varied in accuracy from 50% to 80% (Song et al., 2017; Jadon et al., 2019). Dey et al. (2020) believed that Naïve Bayes is not as accurate as other decision-making algorithms, even with enhanced data cleansing and using AI synergy techniques.

Bayesian Theorem:

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A)x\mathcal{P}(A)}{\mathcal{P}(B)}$$

Support Vector Machine

A support vector machine (SVM) is a decision-making algorithm that can help predict the meaning and context of a sentence, paragraph, and multiple paragraphs. Ren (2019) suggested using an SVM to help increase the accuracy of tweets sentiment by considering the day-of-week effect, which constructs a more reliable and realistic test with stock prices. The day-of-week effect is one of the most well-known financial anomalies; it means that the average return on a Monday is much lower than on the other days of the week. Furthermore, when combined with a stop-loss for Monday's order, the SVM decision algorithm can yield 89% accuracy when predicting stock prices from sentiment analysis of written text about companies on the SSE 50 Index in China (Ren et al., 2019). Attarwala, Dimitrov, and Obeidi (2017) used an SVM to predict the 2012 U.S. presidential election via 100,000 tweet data against the Iowa Electronic Markets (IEM). The IEM is an online futures market for practicing market fluctuations resulting from real-world events such as political outcomes, companies' earnings per share (EPS), and stock returns (Attarwala et al., 2017). Attarwala et al. use of the SVM model correlation output predicted ex-President Obama winning reelection based on Twitter comment sentiments. The study's implied outcome was that Twitter is a valid predictor source for U.S. presidential election outcomes (Attarwala et al., 2017). Zhang, Li, Ye, Li, and Ngai (2015) used an SVM to evaluate the sentiment of two different firms' Twitter feeds and their company stock prices per day and produced a 78% predictive accuracy. Mankar, Hotchandani, Madhawani, Chidrawar, and Lifna (2018) suggested that SVMs are the most efficient and feasible model for predicting stock price movement when it comes to tweet sentiments. However, Akba, Medeni, Guzel, and Askerzade's (2020) study showed that SVMs are not the most effective for predicting cryptocurrency market price fluctuations from sentiment-analyzed tweets. Gunay (2019) found that cryptocurrency during a bull market directly correlates with positive information; however, it is the opposite in a bear market. The use of different methods of algorithms and AI memory and preprocessors is an attempt to enhance SVMs and others when their accuracy is low (Sunitha, Joseph, & Akhil, 2019). An SVM is not a perfect decision-making algorithm. However, it is the best option available for those using sentiment analysis tools.

Neural Network

The development of machines that comprehend and understand emotions is one of the accomplishments of neural networks. The neural network mimics the human brain's biology, the neural cells that use the synopsis to enact brain function (Attarwala et al., 2017). Even today, computer neural networks do not stop with written text comprehension; they can compute hearing and even visual emotions. Furthermore, neural network logic employs a long short-term memory (LSTM) that retains already learned data and continues learning based on what it has in its memory (Sreevidya, Murthy, & Veni, 2020; S. Yildirim, Jothimani, Kavaklioglu, & Basar, 2019). Briefly, recurrent neural networks (RNN), LSTM, and convolutional neural networks (CNN)s are implementations within this study's analysis tool for comprehending sentiment drawn from Twitter and Glassdoor comments. At the root of these machines is a training set with a body of words having a predetermined key-value pair showing whether their meanings are negative, positive, or neutral, along with their synonyms or nearest neighbors to these predefined words mimicking a value closest to a designated sentiment. Neural network algorithms are useful for enhancing predictive modeling for other algorithms such as SVMs (Sreevidya et al., 2020; S. Yildirim et al., 2019). The researcher uses an AI tool to measure sentiment from written text and uses all of the techniques found to produce higher accuracy.

Term frequency-inverse document frequency. How a computer analysis typed text is a complex process of cleaning and analyzing data. A term frequency-inverse document frequency (TF-IDF) produces a key-value pair for each word and its importance in a document (SIET, 2017). The TF-IDF is the beginning of creating a value used for comprehending negative or positive connotations (SIET, 2017). The AI is trained based on a human or, in this case, the researchers showing the program what is recognizable as a negative or positive comment. The AI will also use icons to determine if a sentence is positive or negative; for example, a smiley face has a high probability of positive value (SIET, 2017). The TF-IDF is a cleaning and pre-processing algorithm that helps enhance decision algorithms.

Lemmatization. Cleaning data with multiple processes is a must for accurate computational output. An AI algorithm will need as much information from a sentence as possible to produce an accurate prediction based on the words in a sentence, paragraph, or an entire page and what they mean in context (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). The AI algorithm pre-processing enhancement tool identifies the part of speech intent for each word, sentence, and perceivable context as negative, neutral, or positive sentiment (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). Lemmatization is a useful practice for increasing the accuracy of decision algorithms such as SVM output, but it is not entirely accurate at all times (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). Without the use of lemmatization, however, the AI may confuse the meaning of sentences and provide incorrect output.

Stop words. Some words can confuse a computer and make them mispredict the meaning of a sentence. The algorithms for analyzing parts of speech (POS) use massive computer resources and often miscalculate sentence meaning because of different writing styles (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). For the algorithm to break sentences into useful uniformity, deleting all stop words could increase analysis and output accuracy (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). Stop words such as the, is, at, or which, among many others, are taken out of the analysis (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). However, taking out stop words has the limitation of potentially missing specific context that may lead to inaccurate analysis and context (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). A potential miscalculation can occur when leaving or deleting stops words from a sentence, which is a conundrum combatted by a process called N-Grams.

N-grams. N-grams can increase the accuracy of a computer's ability to comprehend the context of a sentence. N-grams break sentences down into individual words, two-word group grams, three-word group grams, and sentence-grouping grams, with a particular value calculated for each gram (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). The stop words are usually taking out before this process, and the N-Gram values predict the meaning without needing the stop words. Each grams' value is comparable to a set of predefined values offering a means of analysis for determining people, places, sentiment properties, and similarities of context, among other aspects (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). The N-gram process is an analysis that benefits decision algorithms based on increasing output accuracy (Akba et al., 2020; Attarwala et al., 2017; Jadon et al., 2019). However, N-grams can potentially fall into the trap of miscalculations, causing a decision algorithm to predict with bad data at times (Akba et al., 2020; Jadon et al., 2019). These N-gram algorithms still increase the average accuracy of decision algorithms by 10% or more (Akba et al., 2020; Attarwala et al., 2017). The N-gram pre-processing is an important step for analysis and ensuring the best accuracy is probable for sentiment prediction.

Stock prices. Stock prices are an integral measurement for investment decision-making. The measurement of stock prices is done by evaluating each stock's starting day value and comparing the price at the end of the day as higher or lower (Derakhshan & Beigy, 2019; Darena, Petrovsky, Prichystal, & Zizka, 2018; Kumar, Kumar, & Yadav, 2017; Teng-Chieh, Zaeem, & Barber, 2019). Time-series analysis is a valuable measure for stock prices when evaluating minute increments in variances in the price (Darena et al., 2018). The use of time-series analysis is beneficial in forecasting stock prices (Derakhshan & Beigy, 2019; Darena et al., 2018; Kumar et al., 2017; Teng-Chieh et al., 2019). However, this is only necessary for forecasting, which does not fall within the scope of this study. For this study, a simple evaluation of the stock price's initial starting day amount and the end-of-day price will be justifiable through a binary process of price increase or decrease.

Independent Variable: Twitter

The beginning of social media. From the beginning of word of mouth to the creation of the internet, social platforms carry on the necessary marketing process for many businesses. Social media is a prevalent tool that led to the rise of software platforms like Twitter, Facebook, and LinkedIn that hold many different opinions from their 300 million-plus monthly users on products, services, and companies (Asghar, Khan, Ahmad, Qasim, & Khan, 2017; Pak & Paroubek, 2010). Iankova, Davies, Archer-Brown, Marder, and Yau (2018) discussed the different uses of social media from business-to-business (B2B) to business-to-consumer (B2C) marketing models. The outreach of companies advertising goods like commercials bloomed on Facebook and Twitter. The result of this form of marketing was instant customer feedback in real-time, seen by a broad audience that created both a significant opportunity and a problem (Nisar & Yeung, 2018). Even traditional television news outlets began to change because social media platforms gave anyone with a smartphone the ability to publish videos, articles, and opinions viewable on a global scale.

News. The use of news and its psychological effect on traders is an essential topic for behavioral finance researchers. Among the many studies on news sentiment, Kumar et al.'s (2017) paper on Bloomberg news stories and their effect on stock price increases and decreases found a positive correlation. A stock researcher can analyze these news stories for sentiment and relevance to the company using machine-learning algorithms to detect negative or positive content. A conventional method of calculating positive or negative sentiment from findings industry, company, or product-specific news stories and count the number of positive or negative sentiment words in the text (Asghar et al., 2017; Leekha et al., 2017). The effects of these negative and positive words on the investor's decision-making are potential evidence of price spikes from days with news about the company, however less so when the company is not mentioned (Kumar et al., 2017). The amount of news and sentiment seems to have a relationship with investor decision-making.

The amount of time information and its sentiment is saturating in the market, the greater the impact. The length of the news story's effect on the trader's investment decision-making can depend on the amount and type of sentiment (Leekha et al., 2018). The news sentiment effect from positive news can affect stock prices for up to one week (Heston & Sinha, 2017). However, negative news can affect stock returns for up to one quarter (Heston & Sinha, 2017). People and news authors' expression may positively or negatively affect the stock market, which is also called bearish or bullish behavior in stock market terminology (Van de Kauter, Breesh, & Hoste, 2015). Yadav, Kumar, and Kumar (2019a) study on Twitter sentiment analysis on only objective or fact content in news sources proved to have inaccuracies and limitations that led to no relationship and predictability of the Indian NIFTY 50 index price fluctuations.

In contrast, Rani et al.'s (2019) study on the NIFTY 50 index ten-day moving average predictions based on sentiment analysis with one year of 154,000 tweets found a significant correlation. News stories from different outlets may influence the inertia of the Twitter feed. Moreover, the news's effect on a social media platform could enhance the stock prices' sentiment and effect (Zhang et al., 2015). With more social platforms, housing news and other information could lead to increasing the relationship of sentiment and stock price fluctuations.

Social media and breaking news. Social media platforms have enhanced informed decision-making through feedback from many users in the form of negative and positive sentiments. Bollen et al. (2011) mentioned that 19% of their randomly sampled tweets contained positive and negative sentiments that affected specific company stock prices. However, Tamrakar (2016) debunked the theory that stock price fluctuation prediction could come from Twitter, Facebook, Reddit, or YouTube sentiment by analyzing two years of data with 180 firms' stock prices and found no significant relationship. Studying Twitter feedback as an external correlative medium of sentiment is intensely debatable as a useful tool for investor decision-making (Derakhshan & Beigy, 2019). Moreover, investigating internal employee sentiment simultaneously from websites like Glassdoor is not overtly dismissed (Derakhshan & Beigy, 2019). The general reliability of internal and external social feedback sentiments from customers and workers of companies' relationship vis-a-vis enhancing investor decision-making is in question. The investor's use of sentiment information may well sway their decision-making (Asghar et al., 2017). The social media platforms are beneficial for extracting sentiment and using it for investment decision-making.

Social media and investor sentiment. Investor sentiment on social media comes from investors speaking specifically about the stock or performance of any industry or company. Barberis, Shleifer, and Vishny (1998) found investor sentiment to be unreliable when predicting stock prices. The 1998 study instead found that investors underreact to factual information that involved dividend and earning reports and instead reacted to the long history of the firms' exceptional or underwhelming performance. Kaplanski, Levy, Veld, and Veld-Merkoulova (2014) characterized positive sentiment as a relationship of market returns in the private markets. Babu and Kumar (2015) suggested that positive sentiments have a more substantial effect on the NSE index, while negative sentiments had more robust findings in other studies. However, McGurk, Nowak, and Hall (2020) found a correlation between investor sentiment from social media posts and abnormal stock returns, using 3000 random sampled tweets for small capital stocks on the NYSE index, which suggests that small firms have a higher relationship to sentiment than larger firms. The potential power of social media sentiment may be the new word-of-mouth marketing technique that has a relationship to customer decision-making, which affects revenue and perhaps also stock price (Teng-Chieh et al., 2019). Investor sentiment and the sentiment information derivable may or may not be beneficial for investment decision-making.

Trust filters. The influencers of people on social media may be one of the best places to gather information for investment decision-making. Teng-Chieh et al. (2019) discussed the trustworthiness of social media users by filtering their expertise, experience, reputation, and authority. The number of discussions on a topic by such social media users is measurable by their prestige and the number of followers (Teng-Chieh et al., 2019). A strong following or connection to include the number of networks structured around the user is how reputation and authority measurements are derived. Teng-Chieh et al. concluded that influencer posts positively correlate to product sales and stock prices. Teng-Chieh et al. trimmed the users down by 90% in this particular study to only having influencers that have potentially more value to their users (Zaeem & Barber, 2019). The pre-cleaning of users from those who may hold little value to the overall test of sentiment results in a more reliable analysis of sentiment and stock buying and selling decisions. With only 1% of the study data considered reliable for six months from stock prices and influencers, predicting influencer sentiment with stock prices was at 80% (Zaeem & Barber, 2019). An investor may be able to use influencer sentiment information to enhance their investment decision-making.

Individual and Group Accuracy

A single stock relationship to sentiment analysis and its stock price fluctuations compared to a whole index fluctuation is not equivocal. Garcia-Lopez, Batyrshin, and Gelbukh (2018) insisted that sentiment analysis is a useful tool for predicting a stock index price fluctuation, but not a practical choice for individual stocks. Such evidence from Bollen et al.'s (2011) study found that mood states on Twitter data yield 87.6% accuracy when predicting the Dow Jones Industrial Average point fluctuations.

Birbeck and Cliff (2017) found that Twitter sentiment feedback on four separate large technology companies' profiles and their stock price buying and selling automation produced a 54% accuracy rate. The accuracy drops considerably when only evaluating single stock prices. Pagolu, Challa, Panda, and Majhi (2016) conducted a similar analysis with one year's worth of Microsoft tweets and stock prices. Pagolu et al. found an accuracy of 69% in predicting the Microsoft stock price movement. Groß-Klußmann, König, and Ebner (2019) found 190 finance experts' tweets with nine years of data evaluated with sentiment analysis on a few individual stocks. They found a 51% accuracy for predicting future stock prices and investment decision-making. There is a trend from large datasets and predictions averaging around 50% or a little higher, showing signs of the probability of profit having the same chances of something like flipping a fair two-sided coin.

The positive influence on a firm's stock value may also have to do with the current economic state of a country and the perception of the company on social media. Social media affects business performance through four channels: social capital, revealed preferences, social marketing, and corporate social networking (Paniague & Sepana, 2014; Zhang et al., 2015). Many areas from destination planning, consumer confidence, unemployment claims, and more are all derivable and forecasted from social media posts (Paniague & Sepana, 2014; Zhang et al., 2015). There is also evaluating the sentiment from consumers on the different company's social media profiles. Nisar and Yeung (2018) and Usher, Morales, and Dondio (2019) studies found a relationship between consumer sentiment and stock market movements in terms of the relationship between sentiment and the daily opening points with the daily closing points. Zhang et al. (2015) evaluated the sentiment of Twitter feeds from two different firms (CVS and Walgreens) that correlate with variance in the stock's prices each day for six months. The Zhang study findings suggested that there is a stronger correlation from using Twitter feeds than other after-market predictors for stock prices such as Sec 10k evaluations. The amount of information an investor can use to help with investment decision-making is vast; however, the best information is still a well sought-after.

At an international conference on sustainable information engineering technology, the main topic was the use of sentiment analysis on consumer engagement with social media (Saragih, & Girsang, 2017). The conference derived value from sentiment analysis on Facebook and Twitter comments from transportation companies like Uber to see where customer engagements are taking place. It appears that there are more negative comments to include more coming from employees in this industry, and different companies have different values of engagement from different social media platforms (Saragih, & Girsang, 2017). There are areas of the Saragih, and Girsang (2017) study that would create a more substantial quantitative value by adding a Likert-scale to measure the relationship of these companies with customers' and employees' points of view. A combination of consumer and employee analysis could be beneficial information for investor decision-making.

Political Sentiment

Social media has a strong presence in political campaign outcomes and the relationship to stock and index fluctuations. Nisar and Yeung (2018) explored the relationship between related political sentiment from over 60,000 Twitter comments from a U.K. 2016 local election and the FTSE 100 index performance. The study involves measuring FTSE 100 price movements and evaluating correlation via multiple regression analysis by comparing daily moods with daily stock price changes. The findings suggested that there is a correlation between the mood of the public and investment behavior in the short term; however, the relationship was determined to be statistically insignificant (Nisar & Yeung, 2018). Usher et al. (2019) study looking at BREXIT and political sentiment relationship on the FTSE 100 fluctuations found that certain words had a short-term causal effect on index performance.

Twitter post data and sentiment analysis may offer promising analysis results. Oyebode and Orji (2019) analyzed 118,000 Twitter posts about the Nigerian presidential election, trying to predict the outcome one month before the last votes were submitted. The study findings from sentiment analysis found that a single candidate had an 81% chance of winning the election, and the study's prediction was correct. Moreover, Salari, Sedighpour, Vaezinia, and Momtazi (2018) used a similar approach of sentiment analysis from a different social media platform in Iran that found a 97.3% chance of a specific single candidate winning the election, which was later correct as well. The use of sentiment analysis in these studies may not pertain specifically to the stock market; however, the utility of sentiment analysis to predicting relationships is applicable. Overall, these results show promise for using sentiment analysis on Twitter posts and potentially other social media platforms.

Electronic Word of Mouth

The marketing term *word-of-mouth* meaning a person speaking about a product, company, among others, could intentionally or unintentionally influence another people's decision-making. A traditional form of marketing communication between friends and families has evolved into a broad audience around the world because of the Internet (Feng & Liu, 2017). Electronic word of mouth (eWOM) online communication has come in the form of blogs, social media, review platforms, and others (Ramon Saura et al., 2018). The ability to have a more considerable impact through word-of-mouth marketing has evolved with changes in communication platforms to becoming accessible worldwide.

Twitter is such a primary platform for eWOM and a useful evaluative population for specific research questions regarding how comments affect the purchasing of products, services, investments, and even the company's stock price. Xun and Gou (2017) looked into one year of Twitter eWOM feedback sentiment and companies' time-series stock price fluctuation and found a positive correlation. The Xun and Gou data also showed that eWOM has a shorter effect on decision-making than traditional word-of-mouth practices. Large investment firms use eWOM from crowdfunding sites and social media to evaluate seed or any series of funding (Bi, Liu, & Usman, 2018). Bi et al. (2018) suggested that eWOM insights have a relationship to investment decision-making for billions of issued funds. The social media platforms seem to offer more than just entertainment, but also an insight into the possibility of companies' evaluation and success.

The use of sentiment is prevalent for different practitioners, from the consumers to investors using eWOM to garner insight for decision-making. Chu and Sung (2015) studied eWOM on consumers' influence of brands from Twitter posts. Chu and Sung found that customers are more likely to engage in brand discussions when their peers have positive brand attitudes towards a company. Zhao, Wang, Tang, and Zhang (2020) found that peer communication on social media impacts purchasing decisions. Thus, one peer follower's post on a company or product can strongly influence purchasing intentions for many products (Zhao et al., 2020). The influence of peer sentiment may have a strong relationship to decision-making for consumers and investors.

Cryptocurrencies

The new electronic currencies have a place in the investor market. A significant growth cryptocurrency called Bitcoin, reaching the price of $20,000 a coin in 2017, began the shift to a new form of investment currency (Norman & Uulu, 2020). With the rise of Bitcoin came studies testing for sentiment relationships to find ways to predict price changes. Valencia, Gómez-Espinosa, and Valdés-Aguirre (2019) found that sentiment analysis from Twitter comments correlates with predicting the price fluctuations of the major cryptocurrencies. Moreover, Jain, Tripathi, Dwivedi, and Saxena (2018) conducted multiple regression studies on two cryptocurrencies, and the relationship from social media sentiment predicted price change with 59% accuracy. Narman and Uulu (2020) suggested that the relationship between social media sentiment and cryptocurrencies only impacts price change sometimes. However, Inamdar, Bhagtani, Bhatt, and Shetty (2019) found that there is no significant relationship between social media sentiment and the price for any cryptocurrencies. The use of sentiment analysis in the cryptocurrency market is not definitive, just like the stock market. Perhaps other significant investment opportunities such as real estate may fare better.

Real Estate Investments

Sentiment analysis has even reached the commercial and securitized market for real estate decision-making. Ruscheinsky, Lang, and Schäfers (2018) found a constant relationship between news media from 2005 to 2015 by measuring 125,000 news articles and real estate investment trust (REIT) returns and predicting prices for three to four months. Similarly, Hausler, Ruscheinsky, and Lang (2018) found a correlation between news media and individual decision-making based on the perception of sentiment and real estate market movements. However, Braun, Hausler, and Schäfers (2020) suggested that real estate liquidity has a lagging effect after news sentiment is accessible, and investors should use the lag to invest or divest their estate holdings.

Glassdoor

Glassdoor is a software company founded in 2008 that offers employees the ability to have anonymity with ranking and writing opinions about the companies they are currently or have worked for in the past. The companies' ranking is a Likert-scale format allowing the employees to rate their career opportunities, compensation, benefits, culture, values, and the CEO (Chamberlain, 2015). There is also the option to give salary information for the positions held by the employees, which offers other users a sense of the range of potential compensation for the geographical area and title (Chamberlain, 2015). The users can also write positive and negative reviews of the companies with unstructured text that other users view. The large pool of public information from Glassdoor is an excellent database for different hypotheses testing in terms of leadership, employee satisfaction, and benefits' relationships to sentiment analysis.

The chief economist at Glassdoor, Dr. Andrew Chamberlain, publishes many different research papers using the Glassdoor platform data and testing hypotheses. One such research paper published in 2015 analyzed Glassdoor's Best Place to Work companies based on employee feedback and the stock performance of those firms

(Chamberlain, 2015). Chamberlain even used a third-party set of data from Fortune's 100 Best Companies to Work For as an independent variable. The control and comparison variable is the S&P Index 500 performance. The period of measurement was from 2009 to 2014 stock prices (Chamberlain, 2015). The idea that positive feedback from Glassdoor and Fortune's methods of feedback-populated individual companies on their list shows as a portfolio that they outperformed the S&P 500 index by 84% (Fortune) and 115% (Glassdoor; Chamberlain, 2015). The use of the 100 best companies from either source may be a place to help with investment decision-making.

The use of Glassdoor data in the only couple of peer-reviewed articles measures the crowdsourcing reviews and how they affect stock returns. Green, Huang, Wen, and Zhou (2019) suggested that Glassdoor reviews can predict up to three months of stock returns. The study's conclusion is not particularly strong, and the limitation of a single source of data shows a weakness in the lack of independent variables. Sheng (2019) suggested that crowdsourced platforms like Glassdoor can enhance investment decision-making better than traditional sources, with 7 to 9% higher returns. The potential increase for investors looking to enhance their decision-making may need to analyze the sentiment from platforms like Glassdoor.

Internal employee impact. There are many studies on employees receiving stock options and a relationship to changes in business culture, stock prices, and production drive. Aziz, Ansari, and Alam (2019) evaluated the Indian 100 best companies to work with in terms of the employees' view of business culture and relationship to stock prices. The study comes after peer-reviewed studies of United States companies with a similar survey and outcomes of outperforming the market. The best companies to work with surveys measure different business cultural perspectives, including employee turnover and benefits (Aziz et al., 2019). Aziz and Chamberlain suggested that employee feedback may relate to business stock price fluctuations. The scale of the surveys on informational websites like Glassdoor offers a more extensive array of feedback as an opportunity for studying a potential relationship with investment decision-making.

Review of Dependent Variable

Investment decision-making. The types of information an investor uses to execute a decision can be vast or very simple. Individual investors tend to buy or sell based on news, whether positive or negative (Rajdeep, 2020). Moreover, Rubaltelli, Agnoli, and Franchin (2016) suggested that the more a company has a media presence, the more likely the company's stock will yield one-day extreme returns. The type of information and the source of media can be a factor in investment decision-making. The influencers of decision-making come in many different areas from marketing and behavioral finance perspectives from word-of-mouth, professional investment advice, and emotions (Lynderber, 2016; Rubaltelli et al., 2016; Zhou et al., 2018). The use of investment decision-making from AI to insider trading is an extensive program of evaluation in economics and financial services of how these platforms affect a stock market (Rajdeep, 2020). After completing a preliminary investigation of different company information, the use of stock price as an indicator for investment decision-making is a pragmatic approach to buying, selling, or holding stocks (Daniel et al., 2017; Derakhshan & Beigy, 2019; Nisar & Yeung, 2018). Whether small or large, an investment decision is a conclusive commitment, and having the best information to help with the decision is paramount.

Enhancing Investment Decision-making

The breadth of sentiment research from using different data and hypotheses approaches is quite vast. Gao et al. (2020) looked into Google search sentiment and the relationship between 38 different country stock market fluctuations. Gao et al. (2020) found a relationship in predicting all 38 stock market future returns. Che, Zhu, and Li (2020) focused on CEO letters about corporate social responsibility sentiment and their correlation with the company's future financial performance. Li et al. (2017) suggested that some economists believe the information published about a company affects their stock price. Li et al. used 30 large-cap NASDAQ companies' 200 million tweets' sentiment related to the companies' stock price. Li, et al. found 70% accuracy when predicting the future price fluctuations of the 30 NASDAQ companies.

The possibility that a crowd's wisdom can be more influential and correct than single financial professionals' advice is possible. Wu, Ye, Hong, and Li (2020) focused on crowd wisdom and professional analyst recommendation sentiments for investment decision-making to provide relational insights. The professional analyst career field for investment decisions has a lower probability of being correct than a crowd suggestion from social media posts (Wu et al., 2020). The sentiment evaluated in Wu et al. found that both professional and crowd investment suggestions could predict future prices; however, crowd sentiments have a more significant relationship to successful investment decision-making. Moreover, investors only using news article sentiments could be too late when making an investment decision. Hwang and Kim (2019) found that three months of news articles' effect on stock prices have a greater effect on news content than news affecting a stock's price. Zhou et al. (2018) suggested that emotions from sentiment drive 98% of investor decision-making in the Chinese stock market. Zhou et al. evaluated ten million tweets sentiment to the Shanghai Stock Exchange price fluctuations with correlation and Granger causation analysis, producing 64% prediction accuracy. Kesavan, Karthiraman, Ebenezer, and Adhithvan's (2020) sentiment analysis and stock price predictions suggested that sentiment added to an investor's decision-making process is wiser and more profitable. This study should add to the body of research on the subjects of sentiment analysis and investment decision-making.

The United States leads the world in research into how social media affects stock prices and purchasing decisions (Teng-Chieh et al., 2019). The effect of each social media platform and its effect on business performance from many areas such as brand awareness, stock prices, and customer engagement will continue to be an area of future research (Pourkhani et al., 2019). The findings suggest that there are fluctuating relationships and no substantial causation from the studies presented in the literature to date.

Transition

In Section 1, I presented the foundation of the study of a correlation ex post facto study between EMH and successful decision-making measured by stock prices. This section covered the background and statement of the business problem, purpose, nature, and theoretical framework to bound the research. I also state the assumptions and limitations of the study and a review of the possible significance of the study to the practice of business. I concluded with a critical analysis and comprehensive literature review of the EMH theory, measurement tool: sentiment analysis, and the variables of the study: Twitter, Glassdoor, investment decision-making as measured by stock prices. There is also a critical analysis of the EMH theory, independent variables, dependent variables, and alternative theories such as noise trader theory and prospect theory. The section includes the role of the researcher and identification of participants, a discussion of research methodology and design, the role of, and adherence to, the requirements of ethical research, and a review of data collection and analysis techniques.

The data I present in Section 2 includes an overview of the project, the purpose statement, the researcher's role, the participants, and an outline of the research method and design. The further details are the population and sampling method, ethical research, the data collection instruments, data collection techniques, data analysis methods, study validity, and how they pertain to this study. Section 3 includes information about the presentation of the study findings, the application of research findings to professional practice, implications of the study for social change, recommendations for action and future research, reflections, a summary, and conclusions.

Section 2: The Project

Section 2 includes a discussion of the processes used for the execution of the study. I begin with a restatement of the purpose, a discussion of the role of the researcher, and identification of the archival data for this study. I continue with an in-depth discussion of research methodology and design, population and sampling techniques, and ethical research requirements. I close this section with a review of data collection, analysis techniques, and study validity.

Purpose Statement

The purpose of this quantitative ex post facto correlational study was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices. The independent variables were customer comments from Twitter on company profiles, Glassdoor Likert scale ratings, and employee comments on company profiles. The dependent variable was the outcome of investing decisions measured by stock prices. There is no target population of people because I used archival data from U.S. companies listed on the NASDAQ 100 index. The implications for social change might include more effective investment decision-making that may increase funds for retirements, children's education, or other socially desirable uses.

Role of the Researcher

As the researcher, my objective was to answer the research question: Do customer Twitter comments and employee Glassdoor feedback have a relationship with successful investing decisions measured by business stock prices? As an objective observer of a secondary dataset, I have maintained the integrity of the data and have openly discussed any cleaning or manipulation of the data. I was a certified day trader for one year, which sparked an interest in doing this study, including learning about machine learning and data science as a hobby. Prior to the collection of the data, I had obtained Walden University's institutional review board (IRB) approval.

Participants

There were no participants in this study, only archival data with the NASDAQ 100 index companies obtained by secondary dataset from Followthehashtag (http://www.followthehashtag.com/). The companies are large capital firms with active Twitter hashtags and accounts on Glassdoor with over 100 comments on each platform. All of the participants' data were publicly available.

Research Method and Design

The purpose of this quantitative ex post facto correlational study was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions. The study involved secondary data from Followthehashtag's *NASDAQ 100 Companies* dataset and *MIT SMR Culture 500* data that collects and analyzes from publicly accessible Glassdoor employee reviews. The *NASDAQ 100 Companies* dataset had sentiment analysis processes executed from a third-party machine learning algorithm. The *MIT SMR Culture 500* data were already cleaned and analyzed via sentiment analysis from MIT AI algorithms.

Research Method

I have a positivistic mindset as a researcher that influences the method and design for this study. Positivistic researchers seek empirical evidence to suggest truths (Hawang, 2019). I chose a quantitative method for this study because of empirical, publicly available secondary data that is useful for analyzing and answering the research question. The importance of quantitative data is the generalization and capability of testing relationships with independent and dependent variables (Wienclaw, 2013). The understanding of predictors for dependent variables is achievable when using quantitative methods (O'Leary, 2017; Palinkas et al., 2015). If this study had needed interviews and precise nongeneralizable data, a qualitative study could have been more applicable.

Qualitative studies have strengths when specific humanistic areas of data and triangulation approaches are necessary to gain quality results (O'Leary, 2017). If the objective of a study is to gather data from participants with and produce a thematic output for measuring interpretations, then a qualitative study would be useful. However, this is not the focus of this study. The same qualitative processes come with using a mixed-method approach.

A mixed-method approach is the combination of qualitative and quantitative data analysis researchers use to bring both strengths of the methods to a study. Different data types are analyzed depending on the population and available data for a researcher to consider using a mixed-method (Alavi, Archibald, McMaster, Lopez, & Cleary, 2018). The availability of time and money is a factor when using a mixed-method approach that could increase both as a necessity, whereas choosing a single method could be less time-consuming and more cost-effective. Quantitative data collection is the best option for studying relationships with independent and dependent variables (Chirkov & Anderson, 2018; Hwang, 2019). This study encompassed unbiased numerical data that meets the positivist worldview and reasoning for using a quantitative research method.

Research Design

I chose correlation as the design because of the positivistic view and approach to research. A positivistic researcher uses correlation to measure relationships and not causality (Jackson, 2017). In contrast, an experimental design is appropriate for establishing a causation experiment (Bryman & Bell, 2015). The correlation design worked well for this study exploring the relationship between the predictor and independent variables. A researcher using a correlational design can also use longitudinal measurements to analyze data at different times, which was a component of this study (Leedy & Ormrod, 2015). A correlational design and a positivistic view have provided evidence to determine relationships from the study findings.

There are other quantitative designs, such as experimental and quasi-experimental. Quasi-experimental and experimental studies are beneficial for examining the difference between variables when researchers seek cause and effect (Abutabenjeh & Jaradat, 2018). Quasi-experimental and experimental designs also need controllable dependent variables. Since I had no control over the variables, an experimental or quasi-experimental design was not appropriate.

There are three main types of correlational designs, observational, survey, and archival research. The researchers using observational design gather data in a noncontrolled setting from live participants (Boyd, Gove, & Solarino, 2017; Wyman et al., 2015). Researchers using survey design have a preset questionnaire that is later measurable and can add internal validity to the study (Boyd et al., 2017; Wyman et al., 2015). An archival design researcher uses past or secondary data when measuring relationships via data analysis (Boyd et al., 2017; Wyman et al., 2015). All correlation designs have their appropriate need for different research questions and availability of time, cost, and population. I used public secondary data for my study that was already open source and relevant to my research question. Researchers using the archival design or ex post facto design can balance the need for internal and external validity because the data are available to other researchers (Boyd et al., 2017; Handley, Lyles, McCulloch, & Cattamanchi, 2018). The ex post facto design properties can be useful in a quasi-experimental study when evaluating cause and effect, not relationships (Abutabenjeh & Jaradat, 2018). The ex post facto design is beneficial when using publicly available secondary data and effective for correlational analysis and validation testing (Boyd et al., 2017). The data gathered from the secondary dataset were randomly selected; the original researcher gathered the data from *NASDAQ 100 companies'* Twitter feeds, specifically for analysis.

Population and Sampling

Population

There was no population for the study. The archival data came from large capital NASDAQ 100 index companies in a publicly available secondary dataset from the Followthehastag website. The companies were active businesses throughout the chosen timeframe of the study. The firms were publicly owned businesses, not privately owned, with a presence on the NASDAQ stock exchange. The companies' stocks were past the initial public offering stage, which means enough stock pricing data were gatherable to meet the desired timeframe of analysis. I completed a probabilistic random sampling for this study using Microsoft Excel randomizing algorithms to produce such results for analysis.

Sampling

Quantitative studies involve gathering samples in a probabilistic or nonprobabilistic manner. Probabilistic sampling is grabbing participants in a population with the same chances of being selected (Barker & Pistrang, 2016; Palinkas et al., 2015). A researcher determining a probabilistic or nonprobabilistic sample for a study is derivable by time, cost, and population availability (Barker & Pistrang, 2016; Palinkas et al., 2015). The population determination also involves whether the data or population is current and willing to be involved in a study. A nonprobabilistic sampling is appropriate for nonrandomized data, and generalization is not probable from the available data (Barker & Pistrang, 2016; Palinkas et al., 2015). Since this study was a correlational design and provides generalizability, I did not use nonprobabilistic sampling.

There are different probabilistic samples, including simple random sampling, systematic sampling, stratified sampling, and cluster sampling. A simple random sample is having a whole population with an equal chance of being chosen and then using it to analyze (Barker & Pistrang, 2016; Palinkas et al., 2015). A researcher can use a systematic sampling process by choosing participants from a population in a preplanned manner until a sufficient amount of data is obtainable for the study (Barker & Pistrang, 2016; Palinkas et al., 2015). If the sample was to be split up into different groups and chosen randomly, then this method is a stratified sample (Barker & Pistrang, 2016; Palinkas et al., 2015). A cluster sample is selecting random participants in population clusters (Barker & Pistrang, 2016; Palinkas et al., 2015). The sampling method for this quantitative study was probabilistic in the form of simple random sampling.

The secondary dataset in this study consisted of 100 NASDAQ companies, while the full NASDAQ index consists of 3,300 companies' stocks. Simple random sampling from a larger pool of participants and representing a population with a smaller set of data to analyze and generalize is valuable for a study and the researcher's time constraints (Kessler et al., 2015). The value of simple random sampling is the ease of gathering data and representing a larger population (Kessler et al., 2015). A weakness from this sampling method is the requirement to have a full list and equality of all participants in the population, or the generalization factor fails (Barker & Pistrang, 2016; Kessler et al., 2015). Moreover, the necessary amount of data was paramount to study findings and generalizability.

The number of companies needed was anything over 71, including comments from Twitter and Glassdoor reviews for each company. Using G*Power version 3.1.9 software, a power analysis is appropriate to determine the study's appropriate sample size. An a priori power analysis, assuming a medium effect size ($f^2 = .15$), $\alpha = .05$, and two predictor variables, identified that a minimum sample size of 72 participants is required to achieve a power of .80. Increasing the sample size to 146 will increase power to .99. Therefore, I used the secondary dataset *NASDAQ 100 companies* that encompass the target population (see Figure 1). I was not seeking or using small capital or private companies from the NASDAQ dataset or index.

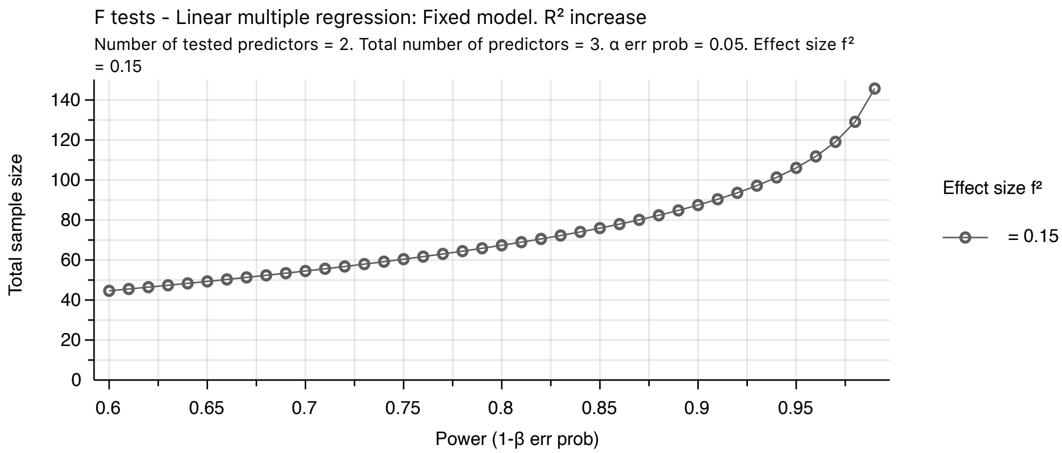

Figure 1. Power as a function of sample size.

Ethical Research

I collected the secondary data for this study that was open to the public for analysis and an open license to use for any purpose. I had completed all necessary IRB form B documentation to enforce the ethical use of the data. The data does not need traditional methods of encryption or protection of personal identifiable information (PII). There was no need to offer incentives or withdrawal procedures, as there were no individuals to provide data for the study. The data are accessible without security firewalls such as a login or password. I protected the company's and Twitter user's identities by not identifying them in this study. The company information is all publicly available information; regardless, I maintained confidentiality by encrypting the data. I will keep the data collected from this study locked in a safe for five years in a thumb drive, and then I will completely wipe the memory destroy the thumb drive with a hammer after five years. The Walden University IRB approval number for this doctoral study is 10-02-20-0759876.

Data Collection Instruments

I used archival data as the data source for this study, and no instruments were necessary. I gathered the archival data from three separate sites Followthehashtag (followthehashtag.com), *MIT SMR Culture 500* (sloanreview.mit.edu/culture500), and Yahoo finance (finance.yahoo.com). Followthehashtag has the *NASDAQ 100 companies* tweets archival dataset available for public use. *MIT SMR Culture 500* is an archive, sentiment analysis, and data visualization website for 550 different companies' Glassdoor post. Yahoo finance website has archives of past stock prices from many different companies.

MIT SMR Culture 500

The *MIT SMR Culture 500* platform was a new online data visualization software created with MIT, Glassdoor, and a company called CultureX. The most recent researcher to use this platform is from a master's thesis from Claremont McKenna College on using Glassdoor data to measure the impact of culture and employee satisfaction (Uyeno, 2020). The Glassdoor review data was from 531 organizations from January 1, 2014, to March 31, 2019. The number of average reviews for all companies was around 2,182, with a median of 1,221 (Sull, Sull, & Chamberlain, 2019). The entire approach for the MIT platform was to evaluate the typed text found on Glassdoor reviews, including the Likert-scale feedback. The type of AI MIT uses is Natural

Language Processing (NLP). MIT SMR provides different evaluation values from the typed text with sentiment analysis and frequency scores in the following nine categories:

- Agility: Employees have the environment to be flexible, nimble, and fast to seize new opportunities or market changes effectively (Sull et al., 2019).
- Collaboration: Employees have success with teamwork practices from different areas of the company.
- Customer: The employees have the customer as the main focus of their everyday practices (Graham, Harvey, Popadak, & Rajgopal, 2017; Sull et al., 2019).
- Diversity: The company promotes and executes diversity and includes it in the workplace without any person's creed at a disadvantage (Sull et al., 2019).
- Execution: Projects managed well, and operational excellence is seen through employee empowerment and adhering to accountable conclusions (Graham et al., 2017; Sull et al., 2019).
- Innovation: The company is a leader in creating cutting edge services or products and ways of working (Graham et al., 2017; Sull et al., 2019).
- Integrity: Employees have a steady presence of honesty and ethical behavior (Graham et al., 2017; Sull et al., 2019).
- Performance: A company rewards employees with many different avenues of appreciation and helps underperforming employees.
- Respect: Employees treat each other with respect (Graham et al., 2017; Sull et al., 2019).

All nine culture 500 categories measured for sentiment are considered a single independent variable for this study. The average mean of the sentiment scores reflects the company's overall sentiment as a whole from Glassdoor reviews.

The data's reliability and validity are reviewed by the MIT Sloan Management peer review committee and created by staff in the MIT and the private industry (Sull et al., 2019). The *MIT SMR Culture 500* used in a 2020 master's thesis at Claremont Mckenna College was proof of academic attention from the data. It was openly accessible for anyone to view.

Monkeylearn

The followthehastag.com NASDAQ 100 dataset was not pre-analyzed for sentiment like the *MIT SMR Culture 500* data. The NASDAQ 100 dataset needed the same sentiment analysis to provide the input necessary for multiple regression analysis. Monkeylearn was an AI online software for early developers or anyone new to AI that provides text sentiment analysis by uploading data to their site (Ramon Saura et al., 2018). The software allows users to select the types of models from Twitter sentiment analysis for financial specific context to conversational context (Rakshitha & Gowrishankar, 2018). The software has an added feature of having the user train the algorithm by reading and manually labeling text with a positive, neutral, or negative sentiment. The feature then uses the user input to output millions of individual Twitter posts sentiment in seconds.

The use of an AI called Monkeylearn for analyzing Twitter data sentiment is found in peer-reviewed studies from the social sciences of evaluating well-being to environmental management sustainability of hotels (Ramon Saura et al., 2018). The use of AI tools was beneficial for those that wish to employ SVM and naïve Bayesian algorithms that are the latest and most accurate AI available to the public. Ramon Saura et al. (2018) used Monkeylearn with online Twitter comments producing a 98.5% accuracy for evaluating correct sentiment from 500,000 reviews. Rakshitha and Gowrishankar's (2018) use of Monkeylearn for determining a person's mental health well-being via 65,994 Twitter tweets came with 86.4% accuracy. The most recent study using Monkeylearn came from a published dissertation about the impact of positive and negative news on the stock performance of sector and market performance, with a rejection of sector and market performance correlating to news and stock performance (Bernardo, 2020).

The Monkeylearn platform was used in peer-reviewed journals for sentiment analysis and used in a 2020 Ph.D. dissertation from Capella University (Bernardo, 2020). A standardized sentiment analysis tool mitigates the biases found in human participation studies (Babu & Kumar, 2015). While using the analysis tool to read emotional responses with standardized words and context of phrasing reduces the need for people to read or synthesizes writings for the sentiment. The analysis tool treats every sentence with the same probability as the next with a constant state of emotional standardization (Babu & Kumar, 2015). The Monkeylearn sentiment analysis tool evaluates large sets of data in seconds for the sentiment. The sentiment output is then dispensed back to the user in a CSV format with each tweet comment and sentiment determination and validation (Ramon Saura et al., 2018; Gowrishankar, 2018). The tool was open source and does not need any special requirements outside of uploading data as a CSV and choose the columns that the system should analyze for the sentiment.

Yahoo Finance

The Yahoo finance intraday data was a valuable asset in peer-reviewed research to include studies involving sentiment analysis. Ranco et al. (2016) used Yahoo Finance data for 100 companies spanning a year-long stock price dataset for each company. Ito, Sakaji, Izumi, Tsubouchi, and Yamashita (2017) used stock market data and board textual data to enhance their machine learning algorithm for rating financial statement sentiment. Moreover, Ho, Damien, Gu, and Konana (2017) used Yahoo Finance message boards content to evaluate sentiment and potential stock market price predictions. The use of Yahoo finance data was prevalent in different types of studies that include stock price metrics and Twitter comments. Thus, providing sufficient evidence for using Yahoo finance as a dependent variable data source for this study.

Data Collection Technique

Is there a statistically significant relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices? The Twitter comments are gatherable from secondary data sources from followthehashtag.com. The secondary dataset was collected from the Twitter API from all *NASDAQ 100 companies* from March 28, 2016, to June 15, 2016 (79 days), consisting of about 1 million tweets. The dates for the secondary data are not significant other than being randomly gathered by followthehashtag.com. The Glassdoor feedback came from *MIT SMR Culture 500*, which has over 513 companies with over 1.2

million comments already analyzed for the sentiment from March 2016 to March 31, 2019. The MIT SMR analyzed data was an excellent source found to be integral from the MIT staff. The only data gathered from the MIT SMR platform came from the *NASDAQ 100 companies'* profiles in the secondary dataset. When gathering the data from the *MIT SMR Culture 500* website, I used the homepage search function to aggregate the company data from the site into an excel spreadsheet. The spreadsheet consists of company names, the number of comments analyzed, and the average sentiment score. The data from followthehastag.com was already formatted with the date, time, context, and added sentiment scores when analyzed by Monkeylearn.com. There is also no need for pilot testing since it was archival data.

When using secondary data or ex post facto, the study's data are available for other researchers to validate and continue future research with the same data (Boyd et al., 2017). The issue with some secondary data is the relevance of the data and how it may not be useful for future researchers with different hypotheses and newer data (Boyd et al., 2017). The secondary data may come from sound sources; however, when using a questionnaire or observational data gathering technique, the researcher controls the data consumption from the beginning, unlike secondary data held first by someone else. There is a level of assumption from secondary data that is not a particular problem from being the first person to gather any data. An advantage with secondary data is the public availability of the data, so other researchers can use it for reliability and validity testing.

Data Analysis

To reaffirm the research question of this quantitative study: What is the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices?

Hypotheses

H_0: There is no statistically significant relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices.

H_1: There is a statistically significant relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices.

As a researcher, I have to consider the types of data and variables when suggesting and using statistical analysis for hypothesis testing (O'Leary, 2017). This study had two independent variables and one dependent variable. The different types of quantitative variables are found in two different general categories continuous and discrete (Bezzina & Saunders, 2014; Mertler & Reinhart, 2017). A continuous variable is a numerical variable between two parameter values: 0 and 1, with no infinite value from fractions, decimals, and whole numbers (Bezzina & Saunders, 2014; Mertler & Reinhart, 2017). Discrete variables have only finite possibilities from two set parameters (Bezzina & Saunders, 2014; Mertler & Reinhart, 2017). This study's two independent variables are continuous variables with the infinite possible values ranging from -1 to 1, including all fractions and decimals. If any score was greater than 0.5 maximum, then the data was represented with a value of 1 for positive sentiment. Any score less than 0.5 maximum is represented as a value of 0, meaning neutral; any score less than 0 is represented as a value of -1, meaning the sentiment is negative (Akba et al., 2020; Attarwala et al., 2017; Birbeck & Cliff, 2017; Derakhshan, & Beigy, 2019). At the same time, this study's dependent variable was also continuous because it is stock prices. The stock prices have a wide range of possible values; however, the starting-day stock price and end-of-day stock price were measurable with a binary construct for this study. An increase of only 1 cent for the stock price means a profit, while a decrease by one cent is a loss represented in the data as 1 being a profit and -1 as a loss (Akba et al., 2020; Attarwala et al., 2017).

Multiple regression analysis and structural equation modeling (SEM) are suitable for studies containing multiple continuous independent variables with a single dependent variable (Henseler, 2017; Kline, 2016). Multiple regression is analyzing all of the independent variables at the same time that enlightens a researcher with the best predictors from the body of data. An SEM is a layer of involved mediators for multiple independent variables, and it tests multiple equations in the same breath (Henseler, 2017; Kline, 2016). The amount of interpretation and deployment setup for the SEM is quite complicated and time-consuming (Henseler, 2017; Kline, 2016). The multiple regression analysis was the most appropriate for this study because there was only one set of independent variables and a single equation derivable for adequate analysis.

The scale of measurements found in quantitative studies is nominal, ordinal, interval, and ratio (Mishra, Pandey, Singh, & Gupta, 2018). A nominal scale is appropriate for assigning numbers to groups such as males equals 1, and females equal two; this is useful when the variable has no fundamental numerical properties (Mishra et al., 2018). An ordinal scale is useful for placement of magnitude, such as rankings from first to last (Bishop & Herron, 2015). The interval scale is an equitable variable populated between two integers and is addressable for quantifying (Mishra et al., 2018). The ratio scale has a definitive definition for zero; the pulse of a human being can is measurable as a ratio (Mishra et al., 2018). The scale of measurement for this study's independent and dependent variables was interval numbers. The use of sentiment analysis outputs a wide range of number properties from -1 to 1 with any possible values between these numbers (Birbeck & Cliff, 2017; Derakhshan, & Beigy, 2019). The dataset had multiple missing values that were preventing the sample from reaching the minimum. I had to input the missing data with bootstrapping, which was the most reliable approach (Ales & Marko, 2017; Sessa & Syed, 2016). All of the raw data in this study is available upon request.

Statistical Analysis Assumptions

When using multiple regression analysis, there are some assumption requirements to consider. The main assumptions are outliers, normality, linearity, heteroscedasticity, heteroscedasticity, and independence of residuals. I used a normal Probability Plot (P-P) of the Regression Standardized Residuals to test the normality, linearity, multicollinearity, independence of residuals, and a scatter plot to test and view potential outliers. The normal P-P is an affective test because of the precision output that is not found in histograms when too little, or too much data is involved (Keith, 2014; Krzywinski & Altman, 2015; Yang & Yuan, 2016). With the main assumptions of multiple regressions discussion, there was a way to inspect for anomalies known as outliers by inspecting the output in a graph.

Outliers can manipulate a relationship test because of the variance imposed on data analysis output (Keith, 2014; Krzywinski & Altman, 2015). Outliers come in many forms: errors, population crossing, typographical mistakes, purposeful and random anomalies (Keith, 2014; Krzywinski & Altman, 2015). I used the normal P-P and scatterplot to review and deal with any outliers by cleaning the outliers out of the data to prevent Type I and Type II errors.

The assumption of normality means all variables are present, and normally distributed data points are sought-after (Keith, 2014; Krzywinski & Altman, 2015). Multiple regressions work best with the normality of data, and to make sure that normality exists, it can be viewed and analyzed with a normal P-P. I also examined for skewness, meaning symmetry of the data, and kurtosis meaning long or short tails scene in the normal P-P output.

The linearity assumption is necessary for multiple regression analysis. If an outcome of multiple regression analysis shows a nonlinear output, the relationship can be skewed and often untrustworthy (Keith, 2014; Krzywinski & Altman, 2015). When a linear relationship is derivable, evaluating the correlation coefficients for random distribution can provide evidence to evaluate homoscedasticity integrity (Keith, 2014; Krzywinski & Altman, 2015). When the heteroscedasticity is profound, the probability of Type 1 errors increases and distorts the study findings (Keith, 2014; Krzywinski &

Altman, 2015; Yang & Yuan, 2016). The multiple regression analysis was essential for evaluating linearity with the independent and dependent variables.

Multicollinearity is the process of testing independent predictor variables for correlation. If a correlation exists between independent variables can make insignificant and unstable estimations of coefficients (Yang & Yuan, 2016; Yoo et al., 2014). When multicollinearity inflates the analysis, the ability to imply the inference or validity of the testing practices is questionable (Yang & Yuan, 2016; Yoo et al., 2014). I tested for multicollinearity by using the variance inflation factor (VIF) to measure the predictors.

The independence of residuals is estimated values from the dependent variable output. A residual is measuring the difference between the estimated and actual dependent value (Keith, 2014; Krzywinski & Altman, 2015). I tested for the independence of residuals by using the normal P-P plot to evaluate random dispersal and constant variance from the residuals (Keith, 2014; Krzywinski & Altman, 2015). The results should be randomly numerical values in ranges of negative and positive $-1 < 0 > 1$ (Keith, 2014; Krzywinski & Altman, 2015). The residuals may present as independent because of potential study design, and understanding the distribution of the data is not necessarily random but conditioned (Ernst & Albers, 2017; Yang & Yuan, 2016).

Multiple Regression Analysis

Analyzing the multiple independent variables allows researchers to view the predictors at the same time. Researchers using a multiple regression need two or more independent variables and at least one dependent variable (Mertler & Reinhart, 2017). I used the multiple regression algorithm with IBM SPSS 25. I imported my CSV or XLS data from the output of the Monkeylearn sentiment analysis. The data cleaning progressed, meaning all of the 1 million tweets had the sentiment signifiers changed from a scale of positive, neutral, and negative sentiment to signifying their interval value: 1 (*positive*), 0 (*neutral*), and -1 (*negative*). The Glassdoor data analyzed from *MIT SMR Culture 500* was labelable with the same scoring of 1 (*positive*), 0 (*neutral*), and -1 (*negative*). The stock market data for each company during the 79 days was evaluated on a daily price basis from the beginning of trading to the end of the trading day. The evaluation of price is accomplishable by noticing if the end of day price is higher than the beginning day price; in this case, the signifier was labelable with a 1, or if there is a lower price, then there is a -1 labeled for the day. The CSV Yahoo finance tables had all available data such as company stock ticker, date, and total day price signifier labeled in each column, and the analysis was commenced, which produced the multiple regression output.

Interpretation of Multiple Regression Analysis

I had tested the multiple regression model to my data for goodness-of-fit via adjusted R^2, F- statistic, and p value with SPSS. The use of R^2 is determining the percentage of variation between models and the data (Keith, 2014; Krzywinski & Altman, 2015). The higher the adjusted $R^{2,}$ the better the fit is between a model and data (Keith, 2014; Krzywinski & Altman, 2015). To reinforce the R^2 output, I used the p value and f-statistic to evaluate the data and model fit assumptions. The p value output definitions come from the output of measurements greater than or less than 0.5. A value that is < 0.5 suggests a model is statistically significant in estimating a dependent variable. A value > 0.5 suggests that changes in the independent variable do not signify a correlational change for the dependent variable (Keith, 2014; Krzywinski & Altman, 2015). A beta (β) is a measurement level for the probability of a Type II error being present. An excellent value a researcher is looking for is 0.20 (Keith, 2014; Krzywinski & Altman, 2015). The outputs of multiple regressions are the beta coefficient, t-statistic, and the p value (Keith, 2014; Krzywinski & Altman, 2015). The testing for the relationship between the independent variables and the dependent variable, the p value tests for the null hypothesis. A p value < 0.05 indicates that the null hypothesis is rejectable, while a value > 0.05 is accepting the null hypothesis (Keith, 2014; Krzywinski & Altman, 2015).

The β coefficients represent the strength between an independent variable and the dependent variable. The β are standardized for each independent variable so that a researcher can differentiate each independent variable strength of relationship to the dependent variable (Keith, 2014; Krzywinski & Altman, 2015). An unstandardized coefficient (B) is an analysis of the dependent variable when independent variables change (Keith, 2014; Krzywinski & Altman, 2015). At the same time, the original units are comparable to the normalized units of the study (Keith, 2014; Krzywinski & Altman, 2015). The standard errors (S.E.) from the standardized and unstandardized coefficients are useful for testing the confidence intervals (CI) and hypothesis. I obtained the output expected with the 95% CI for B, and I am 95% confident the interval contains the population mean.

Study Validity

This quantitative correlation study was a nonexperimental design, and threats to external and internal validity were not applicable. Researchers seek external and internal validity when exploring causal relationships in experimental or quasi-experimental designs. However, threats to statistical conclusion validity inflate the type I error and Type II error rates and are of concern in nonexperimental designs. To reduce the threats of statistical conclusion validity, the analysis instrument selected for this study was reliable. I conducted a pre-study power analysis with G*Power 3 to determine the appropriate sample size and avoided violation of the assumptions for multiple regression analysis.

Statistical Conclusion Validity

A researcher ensuring statistical conclusion validity is obtainable starts with ensuring the sampling methods, statistical tests, and instrument reliability are all correct before executing the study (Anestis, Anestis, Zawilinski, Hopkins, & Lilienfeld, 2014; Stantchev, Colomo-Palacios, Soto-Acosta, & Misra, 2014). If an observable sample or population happens, then the collected data could all be uniform and potentially increase the chances of Type 1 and Type II errors (Anestis et al., 2014; Stantchev, 2014). The use of heterogeneity or different types of samples is a big part of successful regression analysis; however, there are cases of high standard errors and small effect sizes that can impact the test output (Anestis et al., 2014; Stantchev, 2014). The three subject areas I covered in this study are: reliability of the instrument, data assumptions, and controlling Type I and Type II errors.

Reliability of the instrument. I conducted an internal reliability check for the data collection instrument via measurement from the Cronbach's alpha reliability coefficients. Publishers and researchers seek a 0.70 coefficient or higher validation (Da Silva et al., 2013). The completed validation was with the IBM SPSS version 25 software's function for analysis/scale/reliability analysis that produces Cronbach's alpha coefficients. The presented results are in section 3 of this study.

In this study, the instrument for collection and pre-data analysis comes from instruments used in academic institutions and private businesses. The availability of the validity scores is not available at the moment. The prestigious top-level entities that provide context about the tool are Harvard Business Review, The Economist, MIT Sloan, and backed research by the top universities in the country and world (Sull et al., 2019).

Data assumptions. The primary data assumptions are normality, homogeneity of variances, and statistical independence. The normality of data refers to the potential of a normal distribution from the data points gathered (Ernst & Albers, 2017). When using multiple regression, normality is essential when testing assumptions. Homogeneity of variances is having the exact number of variances spread across groups. If there is an imbalance in any group, then false positives can persist and cause incorrect data assumptions. When using an F-test, a researcher can compare variation between the data sets (Ernst & Albers, 2017). Statistical independence is having no relationship between the data sets. I performed a Chi-Square test for evaluating data for statistical independence.

Controlling Type I and Type II errors. To curb bias and increase accuracy, the sample size for collecting and gaining data needs to be appropriate for the study (Hoffman, 2013; Williams, Gomez Grajales, & Kurkiewicz, 2013). Achieving the correct power, confidence levels, and effect size is the next step toward a reliable sample size (Hawkins, Gallacher, & Gammell, 2013; Williams et al., 2013). As a researcher having a larger sample size past the minimum 85%, the study can achieve a better generalization in different populations (Tarhan & Yilmaz, 2014). I had ensured that the followthehashtag.com archive data had the appropriate amount of samples. However, when gathering the data from the other databases, I could not achieve 85% confidence, with only 53 companies data usable.

Statistical Test

A researcher must ensure statistical validity by using an approved statistical test. Assumptions are part of the researcher's process when doing multiple regressions (Mertler & Reinhart, 2017). The data analysis paragraph reviews some of the assumptions addressed when using multiple regression.

The use of bootstrapping is a technique to help with resampling the original dataset encase the first sampling falls short of the requirements (Elvarsson, Taylor, Trenkel, Kupca, & Stefansson, 2014). Bootstrapping is also an excellent substitute for the use of inferential statistics (Elvarsson et al., 2014). The act of bootstrapping uses a randomizer algorithm that duplicates and even forecasts data points from existing data until the desired sample amount is achieved (Elvarsson et al., 2014). I used the bootstrapping resample technique since my sample size was 53 out of the required minimum of 72, as provided by the power analysis found in the sampling paragraph.

Multiple regression assumptions. When conducting a multiple linear regression, the assumptions of normally distributable variables and a linear relationship between independent and dependent variables, measurement reliability, and the variances of errors is homoscedastic is necessary (Casson & Farmer, 2014; Cohen, & Abedallah, 2015). Researchers have several methods for testing assumptions and handling violations. Data cleaning, acceptability of results, exclusion of results, and changing variables are tools available to handle assumption violations (Cohen, & Abedallah, 2015). Researchers that choose not to address violations lessen their ability to infer statistical results in a broader population (Cohen, & Abedallah, 2015).

Data cleaning is a helpful method to improve the distribution of variables by making corrections or mitigating the impact of normal distribution violations when doing a regression analysis (Osborne & Waters, 2002). Using visual plotting graphs is helpful when identifying problems with variable distributions (Osborn & Waters, 2002). I used the IBM SPSS version 25 software to develop and inspect the distribution of plots from the archival data.

Researchers can identify linearity by plotting the dependent variable against predictor variables (Casson & Farmer, 2014). Pearson's correlation coefficient identifies a linear relationship between having a positive relationship when being closer or equal to +1 the opposite when closer or equal to -1 (Mertler & Reinhart, 2017). I used the IBM SPSS version 25 software to measure linearity between successful investment decision-making measured by stock price and each of the predictor variables. Linearity with violations can indicate the regression analysis underestimates the relationship (Osborn & Waters, 2002). All of the nonlinear possibilities from the relationship between variables must be accounted for and considered when stating conclusions (Osborne & Waters, 2002). As a researcher, I accounted for nonlinear possibilities and have found none with the study data.

Overstated or understated results are possible when violating the assumptions regarding the reliability of the measurement (Osborne & Waters, 2002; Cohen, & Abedallah, 2015). Adding additional variables can further complicate reliability and result in less accurate conclusions (Osborne & Waters, 2002). I verified the reliability of measurements by using Cronbach's alpha as a measurement of scale reliability calculated through IBM SPSS software.

Homoscedasticity assumption violations can result in errors and not qual statements across all of the predictor variables in SPSS (Osborne & Waters, 2002). These types of assumptions are observable while reviewing a scatterplot in SPSS. If there is a significant amount of heteroscedasticity, then replacement variables may be necessary to prevent distorted findings (Osborne & Waters, 2002). When completing the test, there were no heteroscedasticity instances found when reviewing the scatterplot.

The external validity of the general population is vital once having an appropriate sample size confirmed by G* Power analysis. With the generalizability of archival data, once gathered, proper external validation is possible for the large population of stock market investors. At the same time, I rely on the population and study findings for external validity (Boyd et al., 2017; Handley et al., 2018). I used the G*power analysis to find an appropriate sample has provided greater reliability and external validity in my study.

Transition and Summary

Section 2 outlined a methodical process for participant selection, research collection, research methods, design, and data analysis. Additional material presented in Section 2 includes a detailed view of the data collection techniques, the purpose statement, the role of the researcher, the population and sampling method, ethical research, data collection techniques, research analysis methods, research reliability, and study validity. The information included in Section 3 contains a presentation of findings, the application of research findings to professional practice, implications of the study for social change, recommendations for action and future research, reflections, a summary, and study conclusions.

Section 3: Application to Professional Practice and Implications for Change

Introduction

The purpose of this quantitative correlational study was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices. The independent variables were customer comments from Twitter on company profiles, Glassdoor Likert scale ratings, and employee comments on company profiles. The dependent variable was successful investing decisions measured by business stock prices. The model could not predict successful investing decisions measured by business stock prices, $F(10, 2993) = .295$, $p = .982$, $R^2 = .001$, meaning the null hypothesis was accepted the alternative hypothesis rejected. The R^2 (.001) value indicated approximately 1/1000% of variations in successful investing decisions measured by business stock prices were accounted for by combined predictor variables. The study findings suggest that there are no significant predictor variables. However, readers should judge the results with caution due to the small sample size and instrument alteration.

Presentation of the Findings

In this section, I discuss the variables' reliability, testing of the assumptions, descriptive statistics, and interpretation of the findings. I conclude with a concise summary. I used the bootstrapping technique to resample the data to address assumption violations because my sample size resulted in 53 complete company data instead of the 77 necessaries for an 85% confidence interval. Small sample sizes often result in low statistical power and imply that statistically significant findings will not reflect a true effect (Button et al., 2013). Where appropriate, I employed bootstrapping 95% confidence intervals using 1,000 samples to address the possible impact of assumption violations.

Reliability of the Variables

The variables used in this study were not instruments of scale. A Cronbach alpha analysis was not necessary for nonscale archival data. A correlation determines the reliability of the data used in this study on the research question and predictor variables reliability. There were no correlations found upon completing the multiple regression analysis. This specific study determines that the preprocessed sentiment analyzed data's reliability may not be reliable enough to predict a relationship between the independent and dependent variables.

Test of Assumptions

I evaluated the assumptions of outliers, normality, linearity, multicollinearity, and independence of residuals. I used bootstrapping with 1,000 samples to focused on the impact of possible assumption violations.

Multicollinearity. I evaluated multicollinearity by viewing the correlation coefficients among the predictor variables. All bivariate correlations were small to medium (see Table 2); therefore, the violations of multicollinearity assumptions were not evident.

Table 2

Correlation Coefficients Among Study Predictor Variables

	P/L	T/Sent.	Agil.	Coll.	Cust.	Div.	Exec.	Inno.	Integ.	Per.	Res.
P/L	1.00	.019	-.006	.014	.004	.006	.004	-.006	-.002	-.010	-.001
T/Sent.	.019	1.00	.142	.051	.047	-.212	.159	-.092	-.113	.067	.038
Agil.	-.006	.142	1.00	-.021	-.150	-.119	.060	.135	-.378	.323	-.518
Coll.	.014	.051	-.021	1.00	.014	.164	-.148	-.641	-.111	.031	-.006
Cust.	.004	.047	-.150	.014	1.00	-.323	-.304	-.166	.028	-.085	-.176
Div.	.006	-.212	-.119	.164	-.323	1.00	-.202	-.305	-.156	-0.35	-.148
Exec.	.004	.159	.060	-.148	-.304	-.202	1.00	.141	-.321	-.127	.109
Inno.	-.006	-.092	.135	-.641	-.166	.305	.141	1.00	-.126	-.200	-.121
Integ.	-.002	.113	-.378	-.111	.028	-.156	-.321	-.126	1.00	-.253	.457
Per.	-.010	.067	.323	.031	-.085	-.035	-.127	-.200	-.253	1.00	-.497
Res.	-.001	.038	-.518	-.006	-.176	-.148	.109	-.121	.457	-.497	1.00

Note. N = 53. Twitter sentiment (T/Sent); Agility (Agil.); Collaboration (Coll.); Customer (Cust.); Diversity (Div.); Executive (Exec.); Innovation (Inno.); Integrity (Integ.); Perception (Per.); Respect (Res.).

Outliers, normality, linearity, homoscedasticity, and independence of residuals. I evaluated outliers, normality, homoscedasticity, and independence of residuals by examining the normal probability plot (P-P) of regression standardized residuals (see Figure 2), a scatterplot of the standardized residuals (see Figure 3), and the histogram of standardized residuals (see Figure 4). I examined linearity with bivariate scatterplots for each independent variable in Figures 5-14). My examination revealed no major violations of outliers, linearity, homoscedasticity, and independence of residuals. However, there was a violation of the assumption of normality.

Figure 2 depicts the normality results of the distribution around the fitted line. My study has a positive skew; however, data distribution showed that most of the points deviated from the line. The study data points' distribution had a wavering deviation from the probability line, which provided evidence that there was a violation of normality assumptions when reviewing Figure 2. I examined the residuals and found there was a pattern to support the assumptions. I used SPSS to compute 1,000 bootstrapping samples to influence violations' possible assumptions, and I reported 95% confidence intervals from the bootstrap samples.

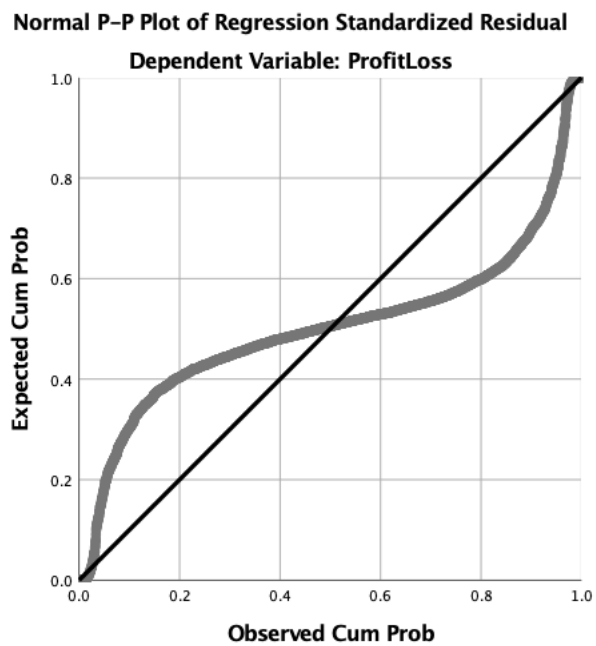

Figure 2. Normal P-P plot of the regression standardized residuals.

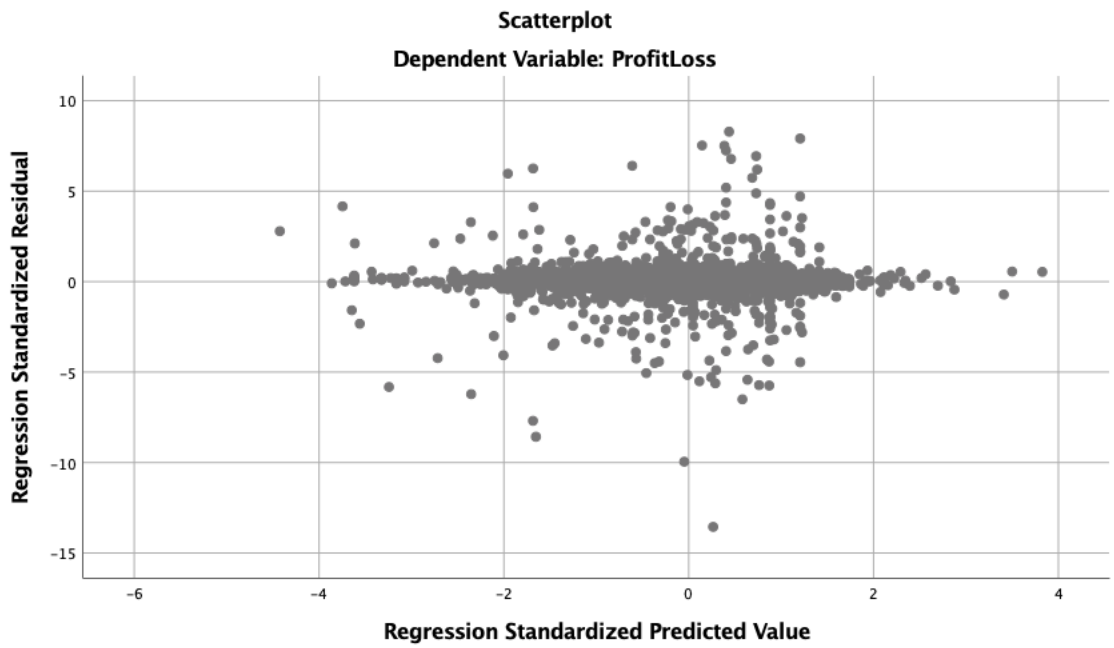

Figure 3. Scatterplot of the standardized residuals.

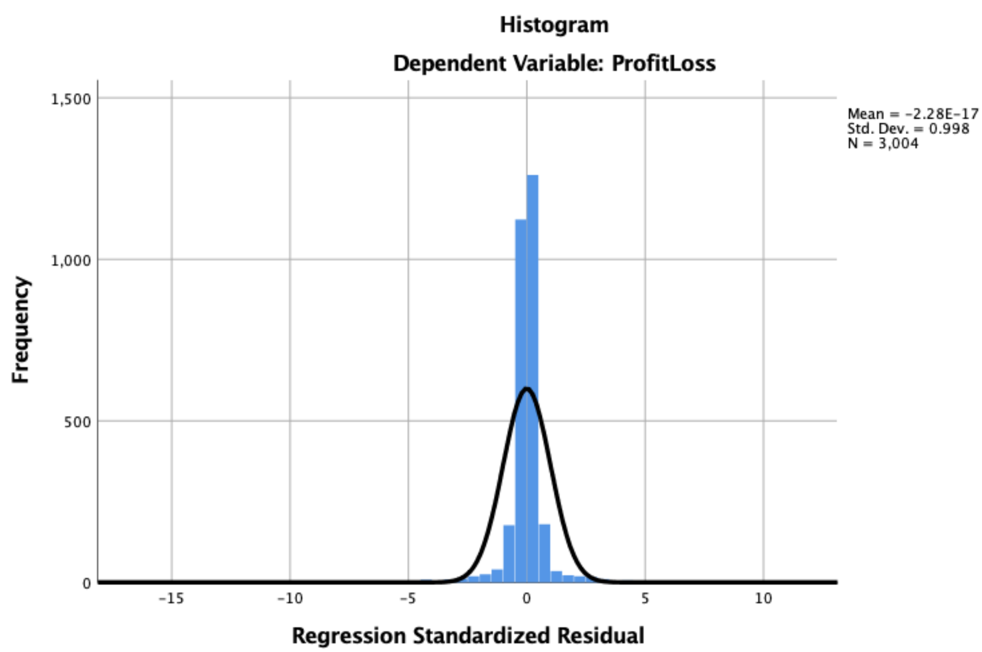

Figure 4. Histogram of the standardized residuals.

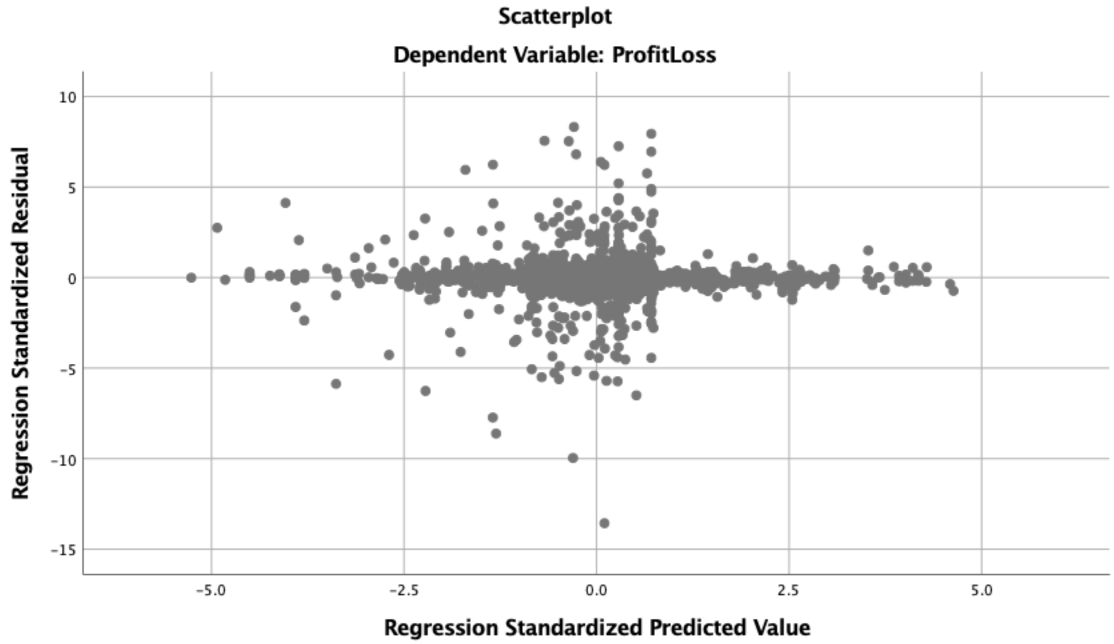

Figure 5. Scatterplot of Twitter sentiment and business stock price profit/loss.

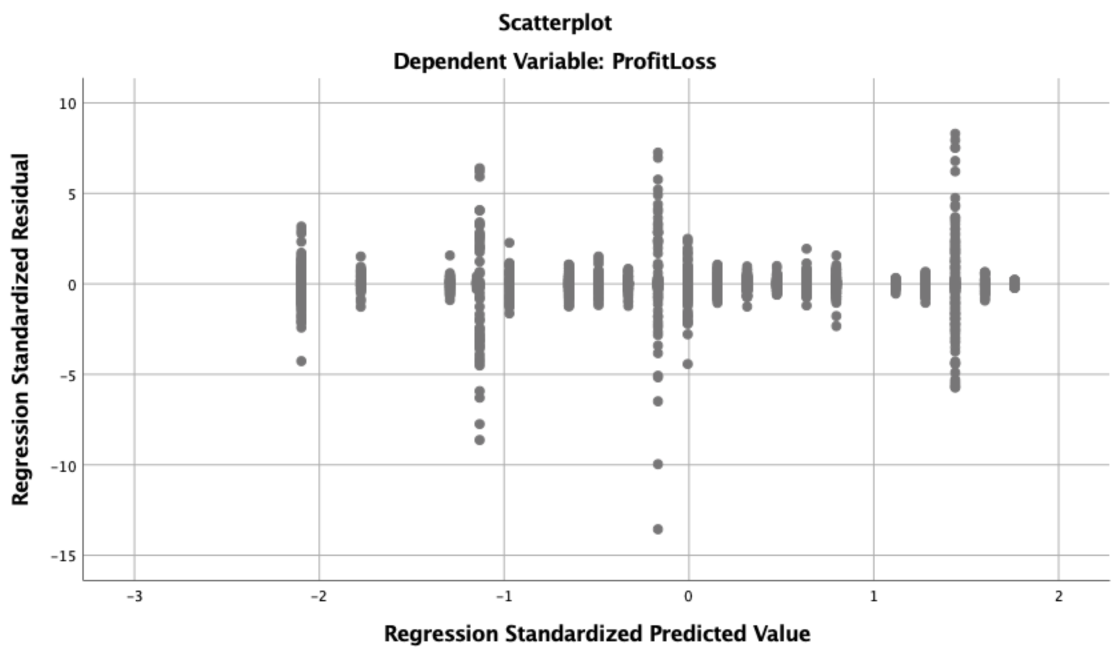

Figure 6. Scatterplot of Glassdoor agility sentiment and business stock price profit/loss.

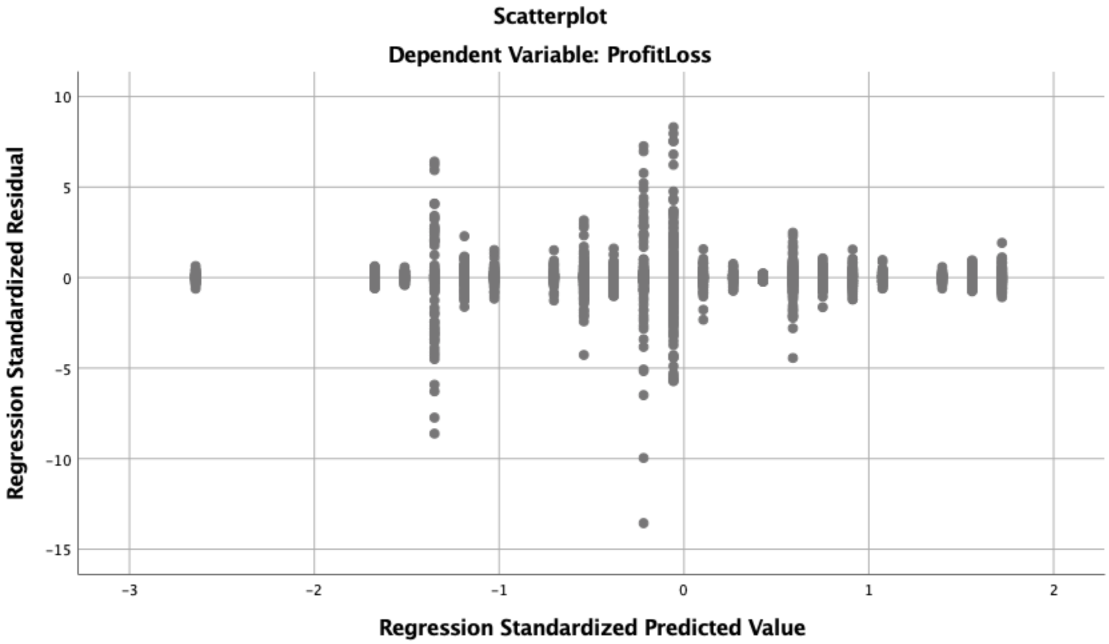

Figure 7. Scatterplot of Glassdoor collaboration sentiment and business stock price profit/loss.

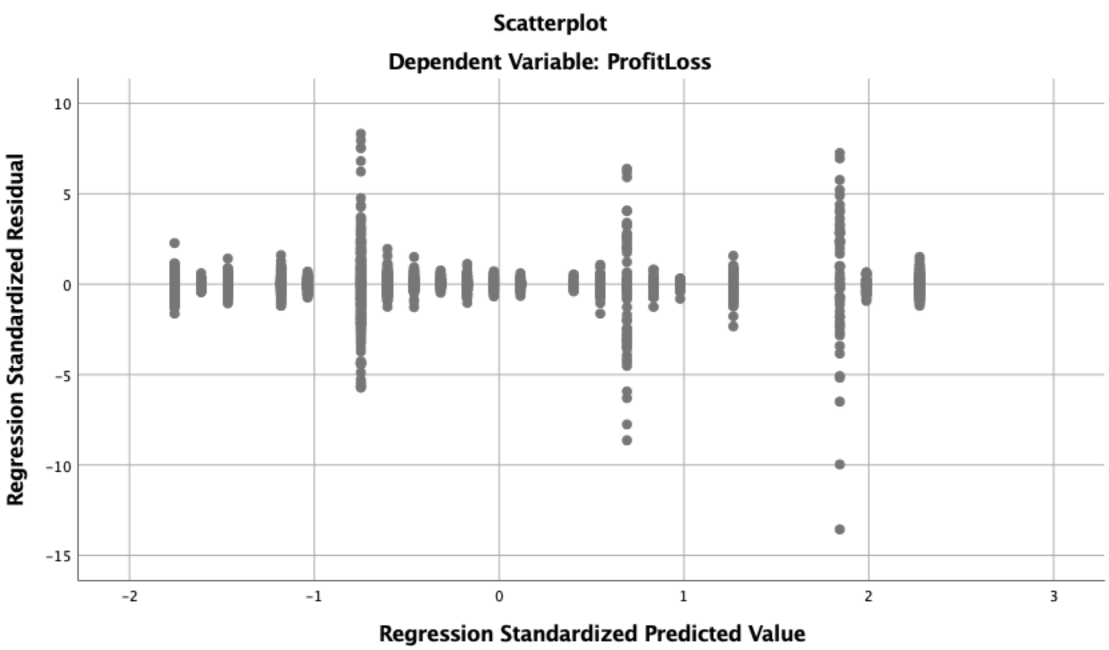

Figure 8. Scatterplot of Glassdoor customer sentiment and business stock price profit/loss.

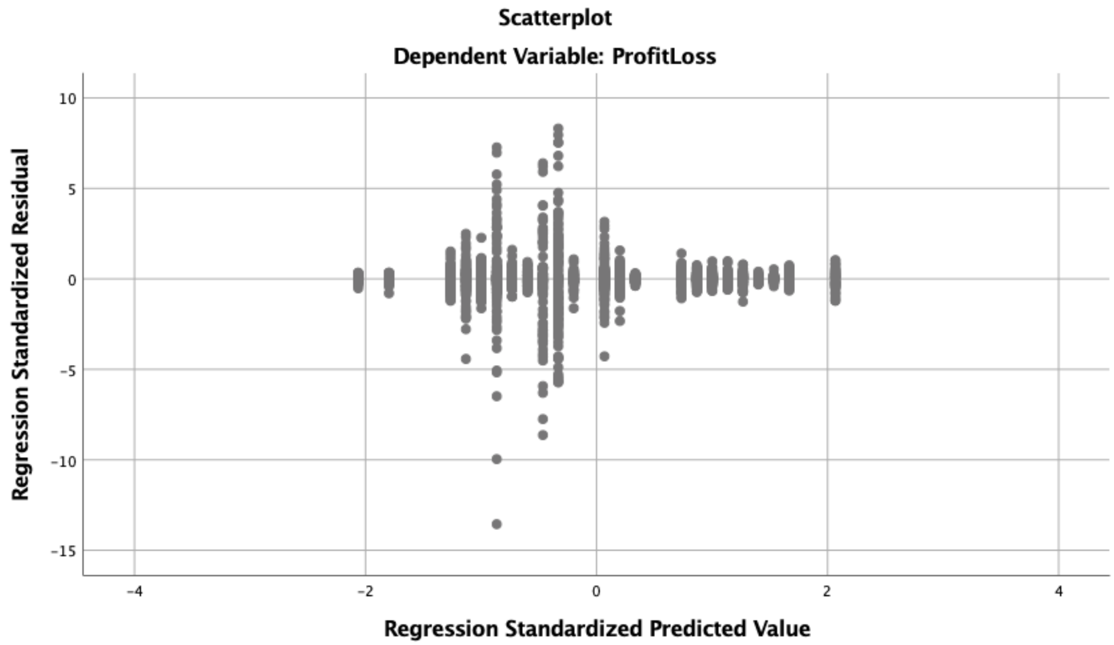

Figure 9. Scatterplot of Glassdoor diversity sentiment and business stock price profit/loss.

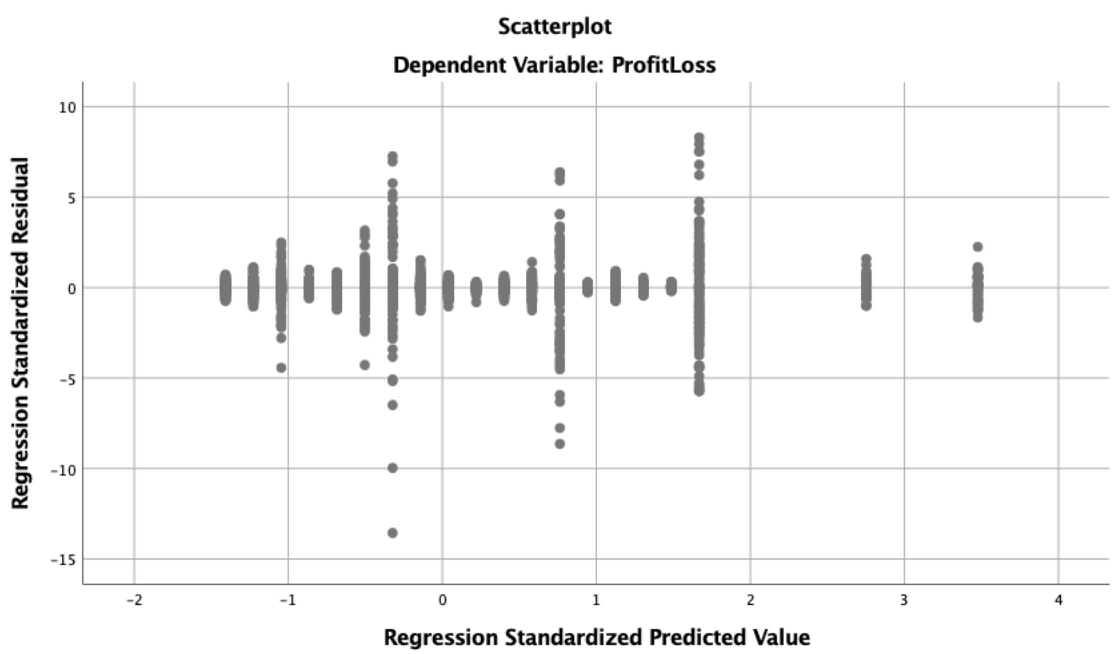

Figure 10. Scatterplot of Glassdoor execution sentiment and business stock price profit/loss.

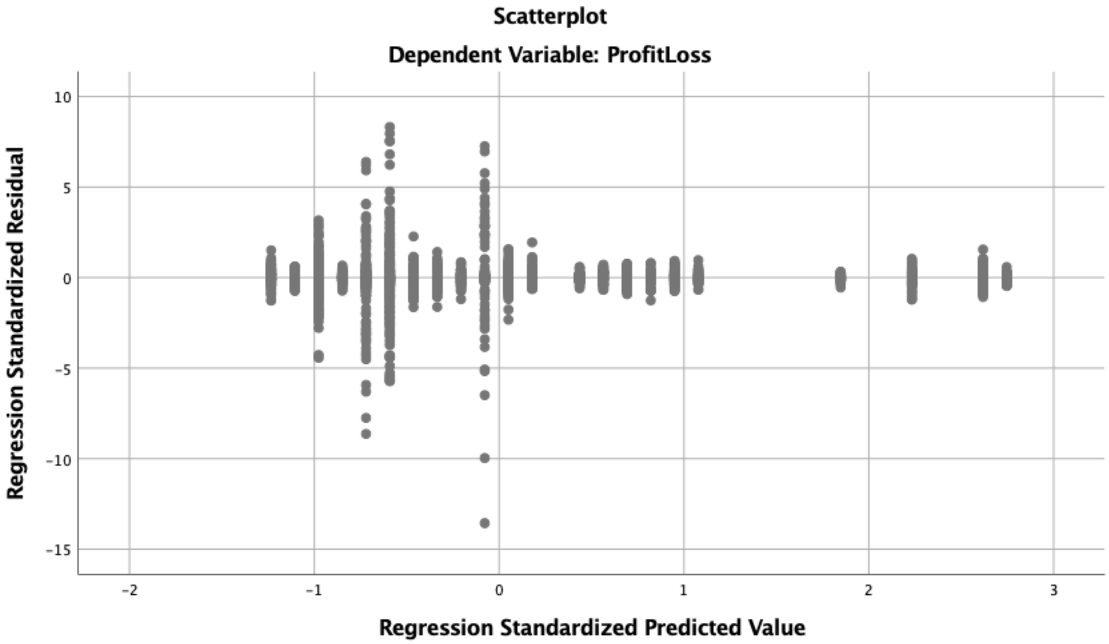

Figure 11. Scatterplot of Glassdoor innovation sentiment and business stock price profit/loss.

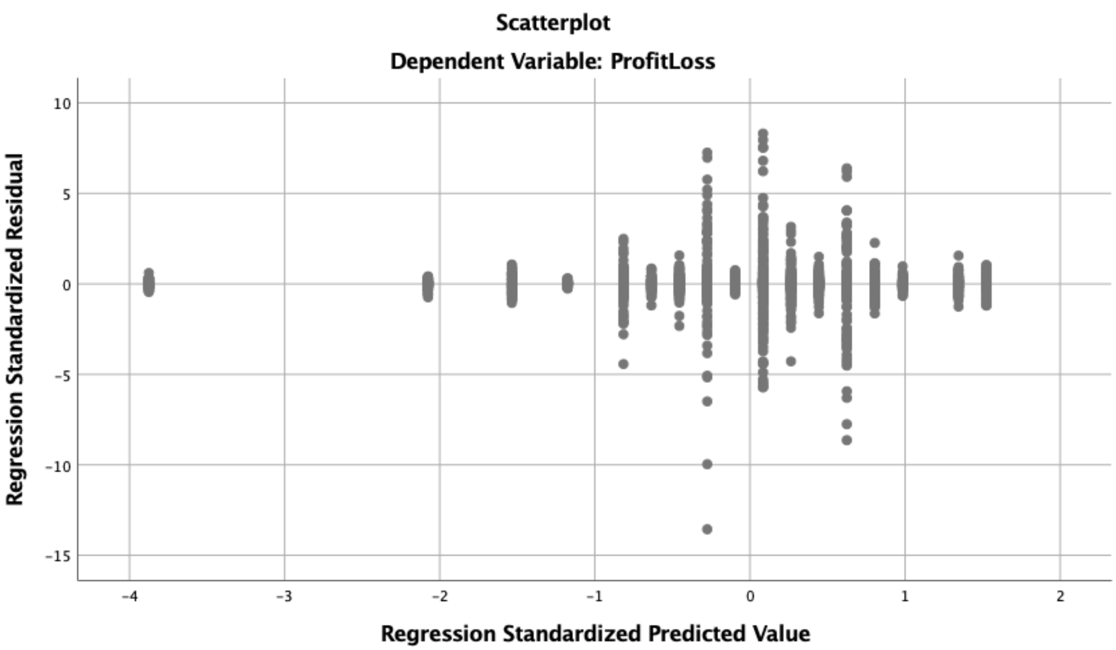

Figure 12. Scatterplot of Glassdoor integrity sentiment and business stock price profit/loss.

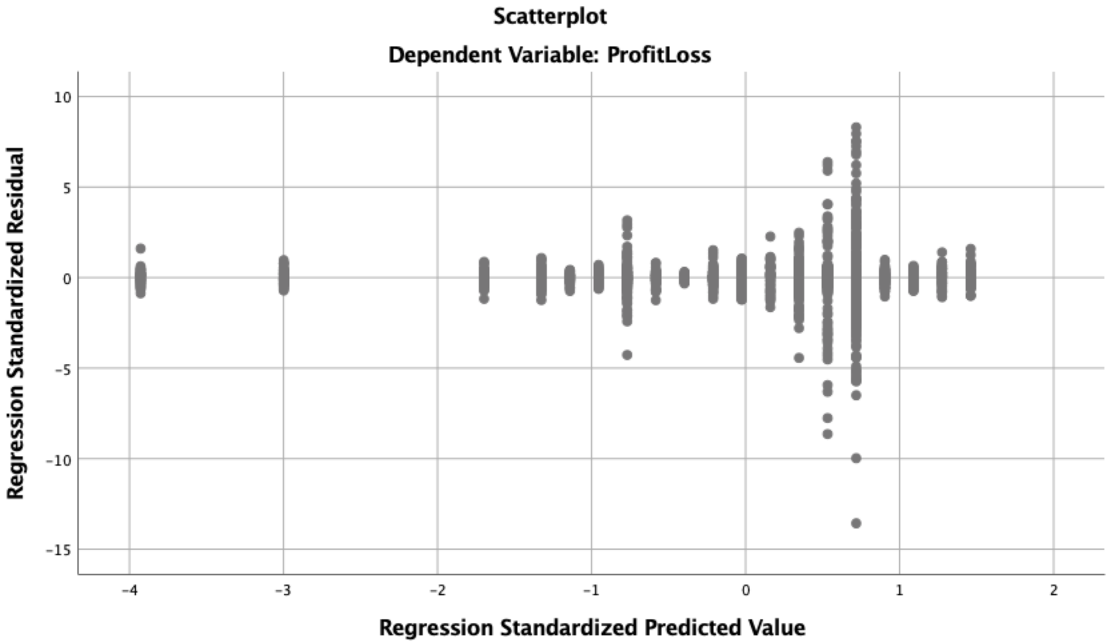

Figure 13. Scatterplot of Glassdoor performance sentiment and business stock price profit/loss.

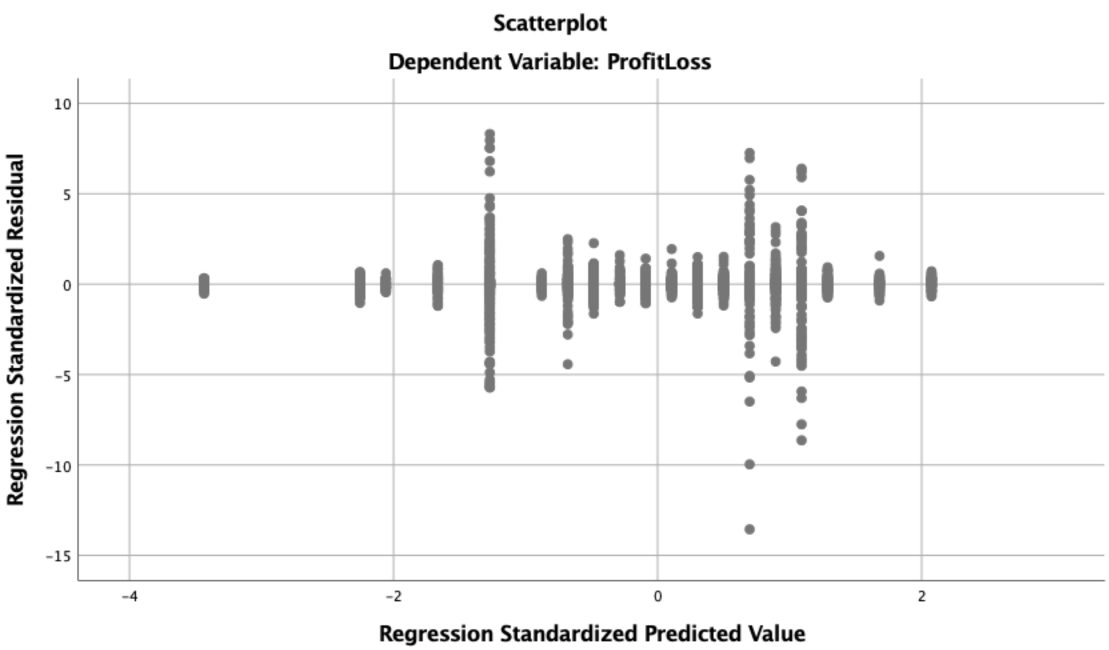

Figure 14. Scatterplot of Glassdoor respect sentiment and business stock price profit/loss.

I found only 53 of the 100 companies on all three of the archival databases. I eliminated 47 companies because their information was not available on either the MITSMR culture 500 or Yahoo finance websites. The small sample size resulted in potential bias in the study. Potential data bias is an error that results from an insufficient number of data available by the target population (Fincham, 2008). Because of the small amount of data present in the data, the reader should judge the results with caution. Table 3 presents descriptive statistics of the independent and dependent variables.

Table 3

Descriptive Statistics for Independent and Dependent Variables

Variable	*M*	*SD*
ProfitLoss	.06	.988
TwitterSent	.47	.375
Agility	-.80	.622
Collaboration	.54	.619
Customer	.42	.695
Diversity	.55	.750
Execution	-.42	.553
Innovation	1.64	.779
Integrity	-.65	.556
Performance	-.41	.538
Respect	-.75	.508

Note. $N = 53$. The *MIT SMR Culture 500* Glassdoor sentiment was measured based on Agility, Collaboration, Customer, Diversity, Execution, Innovation, integrity, Performance, and Respect.

Inferential Results

Of the expected 100 company archival data, I could only use 53 based on the data's availability from all three archival databases. The sample size resulted in significant underpowering of the results. I still used bootstrapping of 1000 samples to evaluate the inferential output. As the number of resampled data sets decreases, the result introduces more variability into the confidence estimation (Haukoos & Lewis, 2005). Having a smaller dataset may have had an inverse effect on the measurement of accuracy.

I used a standard multiple regression analysis to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices. The model was not significantly able to predict successful investing decisions measured by business stock prices, $F(10, 2993) = .295$, $p = .982$, $R^2 = .001$, meaning a rejection of the alternative hypothesis. Table 4 presents a summary of the regression analysis.

Table 4

Multiple Regression Analysis Summary for Predictor Variables

Variable	B	SE B	B 95% Bootstrap CI	*t*	*p*
T/Sent.	.608	.024	[-.367, 1.583]	1.043	.221
Agil.	.004	.001	[-.222, .230]	.036	.971
Coll.	.112	.033	[-.170, .414]	.821	.412
Cust.	.084	.025	[-.237, .404]	.511	.609
Div.	.098	.032	[-.206, .403]	.633	.527
Exec.	.091	.022	[-.205, .387]	.600	.548
Inno.	.095	.032	[-.238, .427]	.558	.577
Integ.	.083	.020	[-.212, .377]	.549	.583
Per.	.026	.006	[-.333, .386]	.144	.886
Res.	.016	.004	[-.375, .408]	.083	.934

Note. N = 53. Bootstrap results are based on 1000 bootstrap samples.

Analysis summary. The purpose of this quantitative correlational study was to examine the relationship between customer Twitter comments, employee Glassdoor feedback, and successful investing decisions measured by business stock prices. I used multiple linear regression to examine the customer Twitter comments, employee Glassdoor feedback to predict successful investing decisions measured by business stock prices. I assessed the assumptions surrounding multiple regression and noted a violation of the assumption of normality. The model, including Twitter comments, employee Glassdoor feedback was not able to significantly predict successful investing decisions measured by business stock prices, $F(10, 2993) = .295$, $p = .982$, $R^2 = .001$. Glassdoor feedback provided useful predictive information about investing decisions measured by business stock prices. The bootstrap computation relies on the sampling distribution when calculating confidence intervals but using extremely small samples will interfere with the computation (Elvarsson et al., 2014). The small original data set was less likely to represent the intended population. Extremely small data sets make it difficult to compute valid confidence intervals (Meeker & Escobar, 2014). This analysis concluded that the predictor variables were not significantly associated with successful investing decisions measured by business stock prices.

Theoretical discussion of findings. A semi-strong EMH is that all public information will reflect in real-time stock prices, meaning gaining a significant profit is not possible when using public information. My findings supported the definition of semi-strong EMH along with the findings of (Nisar & Yeung, 2017; Heston & Sinha, 2017; Tamrakar, 2016, Yadav et al., 2019b), all of whom found that Twitter post sentiment did not correlate with stock price changes nor successful decision-making. However, other studies refute my findings (Bollen et al., 2011, Cwynar et al., 2017, Daniel et al., 2017, Green et al., 2019, McGurk et al., 2020; Usher et al., 2019, 2020

Zhang et al., 2015), and suggest that Twitter post and Glassdoor feedback correlate with firm performance, stock prices changes and successful decision-making. Moreover, this study extended the body of research from using the new *MIT SMR culture 500* sentiment data in the context of hypothesis testing with EMH and investment decision-making. The past peer-reviewed studies that suggest the full conclusion of EMH theory validity are still questionable and an area of future research.

Applications to Professional Practice

This study's findings have applicability to the financial investment professional practice by providing investors with information better to understand sentiment analysis tools' potential misguiding output. The findings of the research, with its limitations, suggest that sentiment analysis tools may not be effective for enhancing investment decision-making. The U.S. Securities and Exchange Commission (SEC) suggests that sentiment tools come with a warning that such tools may cause emotional or risky investment decision-making (U.S. SEC, 2019). The continued research on this subject is something to consider as well; while AI is still enhancing, investors may want to try using sentiment tools sparingly once they are consistently proven to be more useful than traditional investment decision-making practices.

Implications for Social Change

Improving investment decision-making could increase funds for retirements, children's education, or other socially desirable uses. The research may help prevent an investor from using a risky sentiment tool for investment decision-making that may help them prevent loss of capital. Future warnings of new AI investment tools may bring investors to ask for empirical research to help provide evidence of such tools as beneficial and profitable.

Warnings from the SEC about sentiment analysis tools can also reevaluate each year to include the governmental department's most recent stance. The benefits of sentiment analysis to traders may be beneficial in other areas of investment decision-making for evaluating employees' beliefs of diversity, which is a metric found on the MIT SMR 500 culture website. Such cultural sentiment from employees may also help investors change their buying behaviors that may change companies' programs to boost their diversity efforts.

Recommendations for Action

AI sentiment analysis tools relationship to investment decision-making is an on-going debate for the validity of findings. I recommend investors (a) request empirical research before using sentiment analysis investment tools, (b) seek professional financial advice from an accredited institution, (c) and heed warnings from the SEC. The dissemination and acquiring information for sentiment analysis tools may come from the Institute of Electrical and Electronics Engineers (IEEE) articles/conferences and SEC investor bulletins.

Recommendations for Further Research

A limitation of this study was the use of secondary data that may not be accurate. The data used in this study lacked all of the data available for the year 2016 as well. The data was accessible through a Twitter API that is the closest a researcher can get to be the data's first handler. Having the data gathered from the source will decrease the chances of secondary data manipulation.

The next limitation is the amount of data and timeframe of only 79 days in 2016. The amount of data could be increased to a whole year, or longer and account for decision-making based on a week, month, year, or decade from stock prices. The beginning of the data gathering may also start with only the *MIT SMR culture 500* database that may ensure more companies having information on all three databases from Twitter to Yahoo Finance. There are also different types of AI sentiment tools that may help derive more accurate output that may correlate differently with the stock price data.

Another limitation was from using quantitative research that may be too generalized and misses the potential depth of understanding when using a qualitative method. Qualitative studies have strengths when specific humanistic areas of data and triangulation approaches are necessary to gain quality results (O'Leary, 2017). A future research area could be to use qualitative interviews with past and current employees, customers, and professional investors to derive thematic results.

In addition, I would recommend that the data come from other markets and indexes such as the NYSE, DOW, and any publicly accessible market indexes. I had only used NASDAQ 100 companies. Many other companies on the index could be used in a study, or mixing the indexes as a form of random population samples may be useful. I encourage a researcher to gather the appropriate amount of data to fulfill the power analysis range to gain a stronger study output.

Reflections

My Walden University DBA experience was full of ambiguities from committee subjective study input to the publication process's frustrations. A professor told me that my experience is not profound in any way. The road to becoming a doctor is hardly paved and, at times, directionless. However, Walden University's program does offer great information that may, at times, be hidden deep in their databases. Moreover, all of the guidance was there to help me be successful. I also learned that if you do not know what to ask when needing helping with a study, it increases the chances of receiving incorrect information.

However, I have met with more faculty members having a positive effect on me as a student and future doctor. I also used an economic theory that was very interesting that brought to question a lot about what kind of impact or furthering the subject's research would even come from my study. My answer was simple: no one knows if a stock price will go up or down, the past does not necessarily mean predicting the future when looking at the companies past performances. Moreover, the research completed here is a small step to understanding that we may not fully understand if social media sentiment relates to stock prices. However, the AI sentiment tools are constantly being advanced. One day, they may be able to predict stock price outcomes.

Conclusion

The successful decision-making of an investor affects their ability to gain profit leading to philanthropy and increasing the cash flow to a local economy. This study's value to investors was to increase the knowledge of sentiment analysis and investing decisions from customers and employees' posts and feedback. Yadav et al. (2019b) suggested that investors who can decipher sentiment faster than other investors in the active market when making investment decisions can make a more substantial profit than others. However, this study suggests that there is no significant correlation between sentiment analyzed social media content for only 79 days from 2016. Investors must focus on factors that contribute to successful investment decisions and find my studies knowledge to be only a continuation of a conversation, not the end.

References

Abutabenjeh, S., & Jaradat, R. (2018). Clarification of research design, research methods, and research methodology. *Teaching Public Administration, 36*, 237-258. doi:10.1177/0144739418775787

Akba, F., Medeni, I., Guzel, M., & Askerzade, I. (2020). Assessment of iterative semi-supervised feature selection learning for sentiment analyses: Digital currency markets. *2020 IEEE 14th International Conference on Semantic Computing (ICSC),* 459-463. doi:10.1109/ICSC.2020.00088

Alavi, M., Archibald, M., McMaster, R., Lopez, V., & Cleary, M. (2018). Aligning theory and methodology in mixed methods research: Before design theoretical placement. *International Journal of Social Research Methodology, 21*, 527-540. doi:10.1080/13645579.2018.1435016

Alekneviciene, V., Kviedaraitiene, L., & Alekneviciute, E. (2018). Semi-strong form efficiency in the Baltic stock markets under changing economic situation. *Engineering Economics*, *29*(5), 495-506. doi:10.5755/j01.ee.29.5.19083

Allwood, C. (2012). The distinction between qualitative and quantitative research methods is problematic. *Quality and Quantity, 46*, 1417-1429. doi:10.1007/s11135-011-9455-8

Alostad, H., & Davulcu, H. (2017). Directional prediction of stock prices using breaking news on Twitter. *Web Intelligence (2405-6456)*, *15*(1), 1-17. doi:10.3233/WEB-170349

Andrianto, Y., & Mirza, A. (2016). A testing of efficient markets hypothesis in Indonesia stock market. *Procedia - Social and Behavioral Sciences*, *219*, 99-103. doi:10.1016/j.sbspro.2016.04.048

Anestis, M., Anestis, J., Zawilinski, L., Hopkins, T., & Lilienfeld, S. (2014). Equine-related treatments for mental disorders lack empirical support: A systematic review of empirical investigations. *Journal of Clinical Psychology, 70*, 1115-1132. doi:10.1002/jclp.22113

Asghar, M., Khan, A., Ahmad, S., Qasim, M., & Khan, I. (2017). Lexicon-enhanced sentiment analysis framework using a rule-based classification scheme. *PLOS ONE*, *12*(2), 1-22. doi:10.1371/journal.pone.0171649

Attarwala, A., Dimitrov, S., & Obeidi, A. (2017). How efficient is Twitter: Predicting 2012 U.S. presidential elections using support vector machine via Twitter and comparing against Iowa Electronic Markets. *2017 Intelligent Systems Conference (IntelliSys)*, 646-652. doi:10.1109/IntelliSys.2017.8324363

Ayodele, A., Anthony, O., & Adekemi, O. (2018). Semi-strong form of efficiency of Nigerian stock market: An empirical test in the context of input and output index. *International Journal of Financial Research*, *9*(1), 115-120. doi:10.5430/ijfr.v9n1p115

Aziz, T., Ansari, V., & Alam, M. (2019). Corporate culture and stock returns: evidence from Indian companies. *Journal of Strategy & Management*, *12*(4), 464. doi:10.1108/JSMA-01 2019-0009

Babu, A., & Kumar, R. (2015). The impact of sentiments on stock market: A fuzzy logic approach. *The IUP Journal of Applied Finance, 21*(2), 22-33.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics, 49*, 307-343. doi:10.1016/s0304-405x(98)00027-0

Barker, C., & Pistrang, N. (2016). *Research methods in clinical psychology: An introduction for students and practitioners* (3rd ed.) Wiley Online Library. doi:10.1002/9781119154082

Button, K., Ioannidis, J., Mokrysz, C., Nosek, B., Flint, J., Robinson, E., & Munafo, M. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376. doi:10.1038/nrn3475

Bergsma, K., & Tayal, J. (2019). Short interest and lottery stocks. *Financial Management Wiley-Blackwell*, *48*(1), 187-227. doi:10.1111/fima.12220

Bernardo, M. (2020). *The four-week impact of positive and negative news on stock performance in correlation to sector and market performance* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 27741062)

Bi, S., Liu, Z., & Usman, K. (2017). The influence of online information on investing decisions of reward-based crowdfunding. *Journal of Business Research*, *71*, 10-18. doi:10.1016/j.jbusres.2016.10.001

Birbeck, E. & Cliff, D. (2017). Using stock prices as ground truth in Sentiment analysis to generate profitable trading signals. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. doi:10.1109/SSCI.2018.8628841

Bishop, P., & Herron, R. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal of Exercise Science*, *8*, 297-302.

Retrieved from https://digitalcommons.wku.edu/ijes

Bollen, J., Mao, H., & Zeng, X. (2011) Twitter mood predicts the stock market. *IEEE Computer Society. 2*(1) 1-8. doi:10.1109/MIS.2015.25

Bouteska, A. (2019). Some evidence from a principal component approach to measure a new investor sentiment index in the Tunisian stock market. *Managerial Finance*, *46*(3), 401-420. doi:10.1108/MF-11-2018-0570

Boyd, B., Gove, S., & Solarino, A. (2017). Methodological rigor of corporate governance studies: A review and recommendations for future studies. *Corporate Governance: An International Review*, *25*(6), 384 -396. doi:10.1111/corg.12208

Braun, J., Hausler, J., & Schäfers, W. (2020). Artificial intelligence, news sentiment, and property market liquidity. *Journal of Property Investment & Finance*, *38*(4), 309-325. doi:10.1108/JPIF-08-2019-0100

Brutus, S., Aguinis, H., & Wassmer, U. (2013). Self-reported limitations and future directions in scholarly reports: Analysis and recommendations. *Journal of Management*, *39*(1), 48-75. doi:10.1177/0149206312455245

Bryman, A., & Bell, E. (2015). *Business Research Methods*. North York, ON: Oxford University Press.

Cao, H., Ouyang, H., Li, Y., Li, X., & Chen, Y. (2018). The power law characteristics of stock price jump intervals: An Empirical and Computational Experimental Study. *ENTROPY*, *20*(4). doi:10.3390/e20040304

Castellan, C. (2010). Quantitative and qualitative research: A view for clarity. *International Journal of Education, 2*(2), 1-14. doi:10.5296/ije.v2i2.446

Chamberlain, A. (2015, March 11). Does company culture pay off? *Glassdoor*. Retrieved from https://www.glassdoor.com/research/does-company-culture-pay-off-analyzing-stock-performance-of-best-places-to-work-companies/

Che, S., Zhu, W., & Li, X. (2020). Anticipating corporate financial performance from CEO letters utilizing sentiment analysis. *Mathematical Problems in Engineering*, 1-17. doi:10.1155/2020/5609272

Chen, K., Luo, P., Liu, L., & Zhang, W. (2018). News, search and stock co-movement: Investigating information diffusion in the financial market. *Electronic Commerce Research and Applications*, *28*, 159-171. doi:10.1016/j.elerap.2018.01.015

Chinas, M. (2018) *Efficient Market Hypothesis: Weak form efficiency: An examination of weak form efficiency* [Kindle edition]. Retrieved from https://read.amazon.com/ref=kcr_app_surl_cloudreader

Chirkov, V., & Anderson, J. (2018). Statistical positivism versus critical scientific realism A comparison of two paradigms for motivation research: Part 1 A philosophical and empirical analysis of statistical positivism. *Theory & Psychology*, *28*(6), 712-736. doi:10.1177/0959354318804670

Cohen, A., & Abedallah, M. (2015). The mediating role of burnout on the relationship of emotional intelligence and self-efficacy with OCB and performance. *Management Research Review, 38*, 2-28. doi:10.1108/mrr-10-2013-0238

Chu, S., Sung, Y. (2015). Using a consumer socialization framework to understand electronic word-of-mouth (eWOM) group membership among brand followers on Twitter. *Electronic Commerce Research and Applications, 14*(4) (2015) 251-260.

doi:10.1016/j.elerap.2015.04.002

Cakra, Y., & Distiawan Trisedya, B. (2015). Stock price prediction using linear regression based on sentiment analysis. *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS),*147–154. doi:1109/ICACSIS.2015.7415179

Cwynar, A., Cwynar, W., & Pater, R. (2017). Can social media content increase financial market returns? *Organizacija*, *2*, 97. doi:10.1515/orga-2017-0007

Kesavan, M., Karthiraman, J., Ebenezer, T., & Adhithvan, S. (2020). Stock market prediction with historical time series data and sentimental analysis of social media data. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 477-482. doi:10.1109/ICICCS48265.2020.9121121

Krzywinski, M., & Altman, N. (2015). Points of significance: Multiple linear regression. *Nature Methods, 12*, 1103-1104. doi:10.1038/nmeth.3665

Kumari, P. (2019). An empirical analysis of stock price behavior around bonus issue announcement in India. *Journal of Management, 6*(1) 186-196. doi:10.34218/JOM.6.1.2019.020

Da Silva, F., França, A., Suassuna, M., de Sousa Mariz, L., Rossiley, I., de Miranda, R., Espindola, E. (2013). Team building criteria in software projects: A mix-method replicated study. *Information and Software Technology*, *55*, 1316-1340. doi:10.1016/j.infsof.2012.11.006

Daniel, M., Neves, R., & Horta, N. (2017). Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Systems with Applications*, *71,* 111-124. doi:10.1016/j.eswa.2016.11.022

Darena, F., Petrovsky, J., Prichystal, J., & Zizka, J. (2018). Machine learning-based analysis of the association between online texts and stock price movements. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, *21*(61), 95-110. doi:10.4114/intartif.vol21iss61pp95-110

de Aragão Batista, A., Maia, U., & Romero, A. (2018). Stock market under the 2016 Brazilian presidential impeachment: a test in the semi-strong form of the efficient market hypothesis. *Revista Contabilidade & Finanças - USP*, *29*(78), 405-417. doi:10.1590/1808-057x201805560

Derakhshan, A., & Beigy, H. (2019). Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, *85*, 569-578. doi:10.1016/j.engappai.2019.07.002

Dey, S., Wasif, S., Tonmoy, D., Sultana, S., Sarkar, J., & Dey, M. (2020). A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews. *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 217-220. doi:10.1109/IC3A48958.2020.233300

Ding, W., Mazouz, K., & Wang, Q. (2019). Investor sentiment and the cross-section of stock returns: new theory and evidence. *Review of Quantitative Finance and Accounting*, *53(2),* 493-525. doi:10.1007/s11156-018-0756-z

Elvarsson, B., Taylor, L., Trenkel, V., Kupca, V., & Stefansson, G. (2014). A bootstrap method for estimating bias and variance in statistical fisheries modelling frameworks using highly disparate datasets. *African Journal of Marine Science, 36*, 99-110. doi:10.2989/1814232x.2014.897253

Ernst, A., & Albers, C. (2017). Regression assumptions in clinical psychology research practice-A systematic review of common misconceptions. *PeerJ*, *5*, Article e3323. doi:10.7717/peerj.3323

Fama, E. (1965). Tomorrow on the New York Stock Exchange. *Journal of Business*, *38*(3), 285-299. doi:10.1086/294788

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

Feng, J., & Liu, B. (2018). Dynamic impact of online word-of-mouth and advertising on supply chain performance. *International Journal of Environmental Research and Public Health*, *15*(1), 69-75. doi:10.3390/ijerph15010069

Fincham, J. (2008). Response rates and responsiveness for surveys, standards, and the journal. *American Journal of Pharmaceutical Education, 72*, 43. doi:10.5688/aj720243

Gao, Z., Ren, H., & Zhang, B. (2020). Googling investor sentiment around the world. *Journal of Financial & Quantitative Analysis*, *55*(2), 549-580. doi:10.1017/S0022109019000061

Garcia-Lopez, F., Batyrshin, I., & Gelbukh, A. (2018). Analysis of relationships between tweets and stock market trends. *Journal of Intelligent & Fuzzy Systems*, *34*, 3337-3347. doi:10.3233/JIFS-169515

Gillet, R., & Renault, T. (2019). When machines read the web: market efficiency and costly information acquisition at the intraday level. *Finance & Bien Commun*, *40*(2), 7-49. doi:10.3917/fina.402.0007

Graham, J., Harvey, C., Popadak, J., & Rajgopal, S. (2017). *Corporate culture: Evidence from the field* (No. w23255). National Bureau of Economic Research. doi:10.3386/w23255

Green, T., Huang, R., Wen, Q., & Zhou, D. (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics*, *134*(1), 236-251. doi:10.1016/j.jfineco.2019.03.012

Groß-Klußmann, A., König, S., & Ebner, M. (2019). Buzzwords build momentum: Global financial Twitter sentiment and the aggregate stock market. *Expert Systems with Applications*, *136*, 171-186. doi:10.1016/j.eswa.2019.06.027

Gunay, S. (2019). Impact of public information arrivals on cryptocurrency market: A case of Twitter posts on ripple. *East Asian Economic Review (EAER)*, *23*(2), 149-168. doi:10.11644/KIEP.EAER.2019.23.2.359

Guo, L., Vargo, C., Pan, Z., Ding, W., & Ishwar, P. (2016) Big social data analytics in journalism and mass communication. *Journalism & Mass Communication Quarterly*, *93*(2) 332-359. doi:10.1177/1077699016639231

Gyamfi, E., Kyei, K., & Gill, R. (2017). Market efficiency of African stock markets: A meta-analysis. *Journal of Developing Areas*, *51*(4), 69-80. doi:10.1353/jda.2017.0088

Handley, M., Lyles, C., McCulloch, C., & Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health, 39*, 5-25. doi:10.1146/annurev-publhealth-040617-014128

Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: a machine learning approach. *Journal of Property Research*, *35*(4), 344. doi:10.1080/09599916.2018.1551923

Hawaldar, I., Rohit, B., & Pinto, P. (2017). Testing of weak form of efficient market hypothesis: evidence from the Bahrain Bourse. *Investment Management & Financial Innovations*, *14*(2), 376-385. doi:10.21511/imfi.14(2-2).2017.09

Hawkins, D., Gallacher, E., & Gammell, M. (2013). Statistical power, effect size and animal welfare: Recommendations for good practice. *Animal Welfare, 22*, 339-344. doi:10.7120/09627286.22.3.339

Hwang, K. (2019). Positivism versus realism: Two approaches of indigenous psychologies. *Journal of Theoretical and Philosophical Psychology*, *39*(2), 127-129. doi:10.1037/teo0000117

Heston, S., & Sinha, N. (2017). News vs. sentiment: predicting stock returns from news stories. *Financial Analysts Journal*, *73*(3), 67-83. doi:10.2469/faj.v73.n3.3

Ho, C., Damien, P., Gu, B., & Konana, P. (2017). The time-varying nature of social media sentiments in modeling stock returns. *Decision Support Systems*, *101*, 69-81. doi:10.1016/j.dss.2017.06.001

Hwang, E., & Kim, Y. (2019). Interdependency between the stock market and financial news. *2019 IEEE International Conference on Big Data*, 6074-6076. doi:10.1109/BigData47090.2019.9006533

Iankova, S., Davies, I., Archer-Brown, C., Marder, B., & Yau, A. (2019). A comparison of social media marketing between B2B, B2C and mixed business models. *Industrial Marketing Management*, *81*, 169–179. doi:10.1016/j.indmarman.2018.01.001

Ibrahim, N., Wang, X., Bourne, H., (2017). Exploring the effect of user engagement in online brand communities: evidence from Twitter. *Computers in Human Behavior*, *72*, 321-338, doi:10.1016/j.chb.2017.03.005

Ibrahim, N., & Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decision Support Systems*, *121*, 37-50. doi:10.1016/j.dss.2019.03.002

Inamdar, A., Bhagtani, A., Bhatt, S., & Shetty, P. (2019). Predicting cryptocurrency value using sentiment analysis. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 932-934. doi:10.1109/ICCS45141.2019.9065838

Ito, T., Sakaji, H., Izumi, K., Tsubouchi, K., & Yamashita, T. (2017). Development of sentiment indicators using both unlabeled and labeled posts. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1-8. doi:10.1109/SSCI.2017.8280918

Jackson, P. (2017). Causal claims and causal explanation in international studies. *Journal of International Relations and Development, 20*(4), 689-716. doi:10.1057/jird.2016.13

Jadon, P., Bhatia, D., & Mishra, D. (2019). A bigdata approach for sentiment analysis of Twitter data using Naive Bayes and SVM Algorithm. *2019 Sixteenth*

International Conference on Wireless and Optical Communication Networks (WOCN), 1-6. doi:10.1109/WOCN45266.2019.8995109

Jain, A., Tripathi, S., Dwivedi, H., & Saxena, P. (2018). Forecasting price of cryptocurrencies using tweets sentiment analysis. *2018 Eleventh International Conference on Contemporary Computing (IC3),* 1-7. doi:10.1109/IC3.2018.8530659

Jovanovic, F. (2018). A comparison between qualitative and quantitative histories: the example of the efficient market hypothesis. *Journal of Economic Methodology*, *25*(4), 291-310. doi:10.1080/1350178X.2018.1529135

Kaplanski, G., Levy, H., Veld, C. & Veld-Merkoulova, Y. (2014). Do happy people make optimistic investors? *Journal of Financial and Quantitative Analysis*, *50*(1-2), 145-68.

Kessler, L., Barnhart, H., Buckler, A., Choudhury, K., Kondratovich, M., Toledano, A., ... & Sullivan, D. (2015). The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and 99 regulatory submissions. *Statistical Methods in Medical Research, 24*, 9-26. doi:10.1177/0962280214537333

Kim, E., Jeong, Y., Kim, Y., Kang, K., Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science, 42*(6) 763-781. doi:/10.1177/0165551515608733

Kršikapa-Rašajski J., & Rankov-Siniša G. (2016). Testing weak form efficiency on the capital markets in Serbia. *Megatrend Revija*, *13*(1), 265-278.

doi:10.5937/MegRev1601265K

Kumar, A., Kumar, A., & Yadav, R. (2017). Stock price jumps and news sentiment: A case of investor overreaction. *Journal of Prediction Markets*, *11*(2), 20-37. doi:10.5750/jpm.v11i2.1355

Liau, B., Tan, P. (2014). Gaining customer knowledge in low cost airlines through text mining. *Industrial Management & Data Systems, 114*(9) 1344-1359. doi:10.1108/imds-07-2014-0225

Leedy, P., & Ormrod, J. (2015). *Practical research: Planning and design* (11th ed). Essex, England: Pearson Education Unlimited.

Leekha, A., Wadhwa, A., Jain, N., & Wadhwa, M. (2018). Understanding the impact of news on stock market trends using natural language processing and machine learning algorithms. *International Journal of Knowledge-Based Computer Systems*, *6*(2), 23-30. Retrieved from https://www.academia.edu/40508550/Understanding_the_Impact_of_News_on_Stock_Market_Trends_Using_Natural_Language_Processing_and_Machine_Learning_Algorithms

Lekhal, M. & Ahmed, O. (2020). Does the adaptive market hypothesis explain the evolution of emerging markets efficiency? Evidence from the Moroccan financial market. *Heliyon*, *6*(7), 1-12. doi:10.1016/j.heliyon.2020.e04429

Leković, M. (2018). Evidence for and against the validity of efficient market hypothesis. *Economic Themes*, *56*(3), 369-387. doi:10.2478/ethemes-2018-0022

Li, B., Chan, K., Ou, C., & Ruifeng, S. (2017). Discovering public sentiment in social media for predicting stock movement of publicly listed companies. *Information Systems*, *69*, 81-92. doi:10.1016/j.is.2016.10.001

Lydenberg, S. (2016). Integrating systemic risk into modern portfolio theory and practice. *Journal of Applied Corporate Finance*, *28*(2), 56-61. doi:10.1111/jacf.12175

Malkiel, B. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59-82. doi:10.1257/089533003321164958

Malkiel, B., & Fama. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383-417. doi:10.1111/j.1540-6261.1970.tb00518

Manasseh, C., Ozuzu, C., & Ogbuabor, J. (2016). Semi strong form efficiency test of the Nigerian stock market: Evidence from event study analysis of bonus issues. *International Journal of Economics and Financial Issues*, *6*(4), 1474-1490. doi:10.13140/2.1.4941.3448

Mankar, T., Hotchandani, T., Madhawani, M., Chidrawar, A., & Lifna, C. (2018). Stock market prediction based on social sentiments using Machine Learning. *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, 1. doi:10.1109/ICSCET.2018.8537242

Mäntylä, M., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis: A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, 16-32. doi:10.1016/j.cosrev.2017.10.002

McGurk, Z., Nowak, A., & Hall, J. (2020). Stock returns and investor sentiment: textual analysis and social media. *Journal of Economics & Finance*, *44*(3), 458-485.

doi:10.1007/s12197-019-09494-4

Meeker, W., & Escobar, L. (2014). *Statistical methods for reliability data*. Hoboken, NJ: John Wiley & Sons.

Mertler, C., & Reinhart, R. (2017). *Advanced and multivariate statistical methods: Practical application and interpretation* (6th ed.). New York, NY: Routledge; Taylor & Francis Group

Mishra, P., Pandey, C., Singh. U. & Gupta, A. (2018) Scales of measurement and presentation of statistical data. *National Institute of Health*, *21*(4), 419-422. doi:10.4103/aca.ACA_131_18

Narman, H., & Uulu, A. (2020). Impacts of positive and negative comments of social media users to cryptocurrency. *2020 International Conference on Computing, Networking and Communications (ICNC)*, 187-192. doi:10.1109/ICNC47757.2020.9049693

Nwachukqu, J., & Shitta, O. (2015). Testing the weak-form efficiency of stock markets: A comparative study of emerging and industrialized economies. *International Journal of Emerging Markets*, *10*(3), 409-426. doi:10.1108/IJoEM-07-2013-0115

O'Leary, Z. (2017). *The essential guide to doing your research project* (3rd ed.). London, UK: SAGE Publications.

Oyebode, O., & Orji, R. (2019). Social media and sentiment analysis: The Nigeria presidential election 2019. *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 0140-0146. doi:10.1109/IEMCON.2019.8936139

Pagolu, K., Challa, K., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. *International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES).* doi:10.1109/SCOPES.2016.7955659

Palinkas, L., Horwitz, S., Green, C., Wisdom, J., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*, 533-544. doi:10.1007/s10488-013-0528-y

Park, J., & Park, M. (2016). Qualitative versus quantitative research methods: Discovery or justification? *Journal of Marketing Thought*, *3*(1), 1-7. doi:10.15577/jmt.2016.03.01.1

Paraboni, A., Righi, M., Vieira, K., & da Silveira, V. (2018). The relationship between sentiment and risk in financial markets. *BAR - Brazilian Administration Review*, *15*(1), 1-15. doi:10.1590/1807-7692bar2018170055

Polit, D., & Beck, C. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*, *47*, 1451-1458. doi:10.1016/j.ijnurstu.2010.06.004

Pourkhani, A., Abdipour, K., Baher, B., & Moslehpour, M. (2019). The impact of social Media in business growth and performance: A scientometrics analysis. *International Journal of Data and Network Science*, *3*(3), 223. doi:10.5267/j.ijdns.2019.2.003

Rahayu, C. (2017). The influence of information framing towards investors reaction in bullish and bearish market condition: an experiment. *Jurnal Dinamika Manajemen*, *8*(2), 199-207. doi:10.15294/jdm.v8i2.12760

Rajdeep, R. (2020). Past behaviour, financial literacy and investment decision-making process of Individual investors. *International Journal of Emerging Markets*, *15*(6), 1243-1263. doi:10.1108/IJOEM-07-2018-0379

Rakshitha, C., Gowrishankar, S. (2018). Machine learning based analysis of Twitter data to determine a person's mental health intuitive well-being. *International Journal of Applied Engineering Research, 13*(21) 14956-14963.

Ramon Saura, J., Reyes-Menendez, A., & Alvarez-Alonso, C. (2018). Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels. *Sustainability*, *10*(9). doi:10.3390/su10093016

Ranco, G., Bordino, I., Bormetti, G., Caldarelli, G., Lillo, F., & Treccani, M. (2016). Coupling news sentiment with web browsing data improves prediction of intra-day price dynamics. *PLOS ONE*, *11*(1), 1-14. doi:10.1371/journal.pone.0146576

Rani, N., Kaushal, A., & Shakir, M. (2019). Social media and sentiment analysis of Nifty 50 index. *Journal of Prediction Markets*, *13*(1), 50-56. doi:10.5750/jpm.v13i1.1710

Ren, R., Wu, D., & Liu, T. (2019) Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*, *13*(1), 760-770. doi:10.1109/JSYST.2019.2794462

Rubaltelli, E., Agnoli, S., & Franchin, L. (2016). Sensitivity to affective information and investors' evaluation of past performance: An Eye-tracking Study. *Journal of Behavioral Decision-making*, *29*(2/3), 295-306. doi:10.1002/bdm.1885

Ruscheinsky, J., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410-428. doi:10.1108/JPIF-07-2017-0050

Salari, S., Sedighpour, N., Vaezinia, V., & Momtazi, S. (2018). Estimation of 2017 Iran's presidential election using sentiment analysis on social media. *2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 77-82. doi:10.1109/ICSPIS.2018.8700529

Sale, J., Lohfeld, L., Brazil, K. (2015) Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *NCBI, 36*(1), 43-53. doi:10.1023/A:1014301607592

Schneider, K. (2013). Improving data security in small businesses. *Journal of Technology Research, 4*, 1-8. Retrieved from http://www.aabri.com/

Saragih, M., Girsang, A. (2017). Sentiment analysis of customer engagement on social media in transport online. *2017 International Conference on Sustainable Information Engineering and Technology (SIET),* 24-26. doi:10.1109/SIET.2017.8304103

Shahid, M., Jehanzeb, M., Abbas, A., Zubair, A., & Akbar, M. (2019). Predictability of precious metals and adaptive market hypothesis. *International Journal of Emerging Markets*, *15*(5), 1011-1027. doi:10.1108/IJOEM-07-2018-0404

Sheng, J. (2019). Asset pricing in the information age: Employee expectations and stock returns. *SSRN*, 1-63. doi:10.2139/ssrn.3321275

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of bigdata in management research, *International Journal of Production Economics, 191*(1) (2017) 97-112, doi:10.1016/j.ijpe.2017.06.006

Shiller, R. (2003). From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives, 17*(1), 83-104. doi:10.1257/089533003321164967

Sinha, P. C. (2015). Dynamics of noise traders' risk in the NSE and BSE markets. *International Journal of Financial Management*, *5*(4), 30-36. doi:10.21863/ijfm/2015.5.4.021

Song, J., Kim, K., Lee, B., Kim, S., & Youn, H. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. *KSII Transactions on Internet & Information Systems*, *11*(6), 2996-3011. doi:10.3837/tiis.2017.06.011

Sreevidya, P., Murthy, O., & Veni, S. (2020). Sentiment analysis by deep learning approaches. *Telkomnika*, *18*(2), 752-760. doi:10.12928/TELKOMNIKA.v18i2.13912

Stantchev, V., Colomo-Palacios, R., Soto-Acosta, P., & Misra, S. (2014). Learning management systems and cloud file hosting services: A study on students' acceptance. *Computers in Human Behavior, 31*, 612-619. doi:10.1016/j.chb.2013.07.002

Sull, D., Sull, C., & Chamberlain, A. (2019). Measuring Culture in Leading Companies: Introducing the MIT SMR/Glassdoor Culture 500. *MITSloan Management Review. 1-6.* Retrieved from https://sloanreview.mit.edu/projects/measuring-culture-in-leading-companies/

Sunitha, P., Joseph, S., & Akhil, P. (2019). A study on the performance of supervised algorithms for classification in sentiment analysis. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 1351-1356. doi:10.1109/TENCON.2019.8929530

Tamrakar, C. (2016). *Essays on social media and firm financial performance*. (Doctoral dissertation, University of Iowa). doi:10.17077/etd.wuio6qmv

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.

Teng-Chieh H., Zaeem, R., & Barber, K. (2019). It is an equal failing to trust everybody and to trust nobody: Stock price prediction using trust filters and enhanced user sentiment on Twitter. *ACM Transactions on Internet Technology*, *19*(4), 1-20. doi:10.1145/3338855

Uyeno, L. (2020). An empirical analysis of company culture: Using Glassdoor data to measure the impact of culture and employee satisfaction on performance. *CMC Senior Thesis*. Retrieved from https://scholarship.claremont.edu/cmc_theses/2293

Uprichard, E. (2013). Sampling: Bridging probability and nonprobability designs. *International Journal of Social Research Methodology, 16*, 1-11. doi:10.1080/13645579.2011.633391

U.S. Securities and Exchange Commission. (2019, September 18). Investor bulletin: Social sentiment investing tools - think twice before trading based on social media. *SEC Investor Alerts and Bulletins*. Retrieved from https://www.sec.gov/oiea/investor-alerts-and-bulletins/ib_sentimentinvesting

Usher, J., Morales, L., & Dondio, P. (2019). BREXIT: A granger causality of Twitter political polarisation on the FTSE 100 Index and the pound. *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE),* 51-54. doi:10.1109/AIKE.2019.00017

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Entropy*, *21*(6), 589. doi:10.3390/e21060589

Van de Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications, 42*, 4999-5010. doi:10.1016/j.eswa.2015.02.007

Wienclaw, R. (2013). Quantitative and qualitative analysis. *Research Starters: Sociology (Online Edition)*. Retrieved from www.ebsco.com/

Williams, M., Gomez Grajales, C., & Kurkiewicz, D. (2013). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research & Evaluation, 18*(11), 1-14. Retrieved from http://pareonline.net/

Worthington, A. (2013). Financial literacy and financial literacy programmes in Australia. *Journal of Financial Services Marketing, 18*, 227-240. doi:10.1057/fsm.2013.18

Wu, X., Ye, Q., Hong, H., & Li, Y. (2020). Stock selection model based on machine learning with wisdom of experts and crowds. *IEEE Intelligent Systems*, *35*(2), 54-64. doi:10.1109/MIS.2020.2973626

Wyman, P., Henry, D., Knoblauch, S., Brown, C., Wyman, P., & Brown, C. (2015). Designs for testing group-based interventions with limited numbers of social units: The dynamic wait-listed and regression point displacement designs. *Prevention Science*, *16*(7), 956-966. doi:10.1007/s11121-014-0535-6

Xun, J., & Guo, B. (2017). Twitter as customer's eWOM: an empirical study on their impact on firm financial performance. *Internet Research*, *27*(5), 1014-1038. doi:10.1108/IntR-07-2016-0223

Yadav, R., Kumar, A., & Kumar, A. (2019a). Event-based sentiment analysis on futures trading. *Journal of Prediction Markets*, *13*(1), 57-81. doi:10.5750/jpm.v13i1.1731

Yadav, R., Kumar, A., & Kumar, A. (2019b). News-based supervised sentiment analysis for prediction of futures buying behaviour. *IIMB Management Review (Elsevier Science)*, *31*(2), 157-166. doi:10.1016/j.iimb.2019.03.006

Yang, M., & Yuan, K. (2016). Robust methods for moderation analysis with a two-level regression model. *Multivariate Behavioral Research, 51*, 757-771. doi:10.1080/00273171.2016.1235965

Yildirim, A., Sküdarlı, S., & Zgür, A. (2016). Identifying topics in microblogs using Wikipedia, *PLOS ONE, 11*(3) 1-20. doi:10.1371/journal

Yildirim, S., Jothimani, D., Kavaklioglu, C., & Basar, A. (2019). Deep learning approaches for sentiment analysis on financial microblog dataset. *2019 IEEE International Conference on Big Data,* 5581-5584. doi:10.1109/BigData47090.2019.9006056

Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q., & Lillard, J. (2014). A study of effects of multicollinearity in the multivariable analysis. *International Journal of Applied Science and Technology, 4*, 9-19. Retrieved from http://www.ijastnet.com/

Yoosin, K., Rahul, D., Jie, Z., Seung, J. (2016). Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. *Online Information Review*, *40*(1), 42-61. doi:10.1108/OIR-03-2015-0068

Zhang, W., Li, C., Ye, Y., Li, W., & Ngai, E. (2015). Dynamic business network analysis for correlated stock price movement prediction. *IEEE Intelligent Systems*, *30*(2), 26-33. doi:10.1109/MIS.2015.25

Zhao, Y., Wang, L., Tang, H., & Zhang, Y. (2020). Electronic word-of-mouth and consumer purchase intentions in social e-commerce. *Electronic Commerce Research and Applications*, *41*. doi:10.1016/j.elerap.2020.100980

Zhou, Z., Xu, K., & Zhao, J. (2018). Tales of emotion and stock in China: Volatility, causality and prediction. *World Wide Web*, *21*(4), 1093-1116. doi:10.1007/s11280-017-0495-4