Twitter Mood in Predicting the Stock Market using Sentiment Analysis

Matthew Su ; 2019; Word Count: 4540

***Abstract*—In behavioral economics, emotions, especially when regarded as public mood, may affect future decision-making in individuals. In this paper, I determine whether public mood in the form of average sentiment derived from Twitter feeds correlated to the value of the Dow Jones Industrial Average (DJIA) as a function of time. Using sentiment mood analysis techniques such as TextBlob, I receive a sentiment polarity score between -1 and 1, where the scores indicated negative and positive sentiment, respectively. A Granger causality analysis and a Recurrent Neural Network (RNN) in the form of a Long Short Term Memory Model (LSTM) are then used to test the hypothesis that that public mood in the form of Twitter sentiment is predictive of DJIA value stock returns. My results showed that the twitter sentiments derived from a Twitter API were not predictive of future stock changes with statistical values greater than an alpha value of 0.01. In addition, directional accuracy of DJIA predictions resulted in a Root Mean Square Error (RMSE) value of 1227.6668584872302.**

I. Introduction

With the development of social media, exposure to public opinion has become inevitable. The awareness and accessibility of social media data has grown significantly. Social media is an online platform for sharing emotions publicly about any subject, thus, affecting public opinion [1]. In recent years, Twitter—a microblogging service—has become an interesting and viable source for researchers to use as a source of data. Providing information in real time, Twitter connects users and informs them about subjects that might strike future interest or affect future decision-making. Microblogging platforms are used by many people to express and present their opinions are particular subjects; thus, Twitter (and other forms of social media) can be determined to be a valuable source of people's opinions [2]. The public mood is affected by over 200 million users on Twitter, and motives differ from user to user [2]. Cumulatively, Twitter users post more than 400 million tweets a day, providing researchers with useful information on public opinion.

Although the maximum word capacity of tweets are 280 characters per "tweet," the accumulation of the hundreds of millions of tweets posted on the internet in any timeframe must be examined in order to present an accurate depiction of sentiment and public mood [1]. A single tweet displayed by a user may not be perceived as a valuable source of information on public sentiment; however, the aggregation of tweets—groups of tweets collected into a cluster—when analyzed, may provide insight on sentiment and public opinion, which may be used for market trading [2]. If aggregated tweets about specific companies are utilized and correlated with economic indicators referring to the financial market and stock exchange, there may be a relationship between consumer and market sentiment. This paper aims to investigate the polarity of the correlation between public sentiment, which is expressed in large-scale datasets of daily Twitter feed, and trends in the stock market, which are expressed as the closing price of a particular market index fund. Thus, determining Twitter feeds effectiveness can be used to predict the stock market.

II. Literature review

A. Efficient Market Hypothesis

Prior research [3], [4], done on stock market prediction, was oriented around a random walk theory and the Efficient Market Hypothesis. The Efficient Market Hypothesis (EMH) is the theory that stock market prices are primarily driven by live information, e.g., live information such as news rather than market sentiment in the past. [5] Therefore, EMH suggests that the stock market and the prices associated with companies in the stock market follow a random walk structure, thus, null with predictions with more than a 50% accuracy due to live data being rendered as useless, unable to predict any future market movements [5]. For example, [5] demonstrates that since the flow of information is uninterrupted and information immediately reflects the stock prices, then the following day's price change will reflect only following corresponding news and will be independent of the price changes today. However, 2 problems coincide with the EMH. First, several studies have shown that stock market prices fail to follow a random walk pattern, challenging the EMH in that the stock market can be predicted to some extent [6]. Second, early indicators from social media (Twitter, blog posts, Facebook, etc.) may be used as economic indicators, which ultimately could be the case for the stock market as well. Therefore, the development of technology that provides real time data may lead to a correlation for trends occurring now. For instance, [7] explains a study in which recent online data activity correlated with and predicted actual consumer purchase decisions at online retailer Amazon. With the development of social media and technology, I can safely challenge the fundamental idea of the EMH as the base assumption in my methodology. However, my challenge does not refute the idea that real time data affects the stock market; rather, the challenge contests that previous data may affect the stock market as well [5].

B. Mood Sentiment

Public mood and sentiment may also play an important role; previous research reveals that in addition to empirical information, emotions also play a significant role in consumer decision-making [8]. Personality and attitude from the appeal of outside information can be helpful in an investor's decision for investment strategies, particularly towards individual companies of interest [8],[9]. In behavioral finance, emotions and mood have been proven to significantly drive finical decisions [10]. Therefore, one can reasonably assume that moods and the personality of the public can fluctuate stock market trends just as significant as market news can. Prior research done by [11] proves this assumption by extracting social media sentiment moods and finding a correlation between prior sentiment and Dow Jones Industrial Average (DJIA) values.

In order to study how public emotions/moods influence the stock markets as technology and information abundance progresses over time, a reliable, measurable, and relevant assessment of the public mood in terms of a timeline is essential to construct an appropriate model for stock market prediction. Applicable surveys and daily live posts of public mood over random samples for representation of the population may be costly in money and time as demonstrated by Gallup's opinion polls and certain indices [12].

Previous research [13], [14] utilized a technical analysis and observation of public mood or sentiment from soccer games results and weather conditions, respectively. However, a lack of accuracy in the data is an impediment to the minuscule level to which independent variables are assumed to be correlated with the public's mood sentiment. Since 2005, significance has been established in using different sentiment techniques that utilize mood indicators from the public directly from multiple social media platforms like blogs [15] and more specifically large-scale Twitter feeds [16]. The aggregate of hundreds of millions of tweets posted and presented on Twitter in any timeframe may provide accurate representations of public mood and their sentiment [2]. As shown in [17], the use of online consumer data has led to the development of real time sentiment-tracking indicators, thus further shown by several studies supporting a dismissal of the EMH [18],[19].

C. Gap

Thus, the gap is shown due to the lack of research on how social media may influence the stock market. Due to the rapid development of social media platforms, it is important to address its potential effects on our economy.

With the abundance of data presented as opinions on micro-blogging platforms and the progression of computer programming techniques in order to enhance predictions in the stock market, the question arises: *To what extent do Twitter and the sentiment presented with it correlate to the trends occurring in the stock market; specifically the Dow Jones Industrial Average (DJIA) values using* *sentiment-analysis techniques in Python?* In this paper, I investigate the hypothesis that public mood in the form of Twitter sentiment is predictive of changes in DJIA values.

III. Methods

The Dow Jones Industrial Average (DJIA) served as a suitable model and subject for this study because the index fund accomplished several tasks for the purpose of this paper. The DJIA is a stock market index that indicates the stock performance of 30 large companies listed on several stock exchanges in the United States (New York Stock Exchange (NYSE), NASDAQ). Having an average trade volume of 508,876,935 (as of 4/19/2020) and an estimated market cap of $8.33 trillion (Dec. 2019), the DJIA is composed of a large portion of the U.S. economy and stock market [20], [21]. In addition, the DJIA is a great representation of the stock market in terms of public awareness as the DJIA includes a multitude of large well-known companies: Apple Inc, McDonald's, Microsoft, etc.. Being that the DJIA includes large companies, the DJIA is subject to sentiment more than other market index funds such as the NASDAQ-1000 and S&P 500. While the other market funds include more companies in their indexes, they lack in company size and public awareness as they are primarily composed of small-cap companies. Therefore, DJIA, with its large public awareness and large composition of the stock market, was an ideal subject to compare to mood sentiment in this study.

A. Extracting Market Data

The stock data for DJIA was collected daily in a one month period between March 18, 2020, and April 22, 2020. Using Yahoo! Finance, the values (Open, High, Low, Close*, Adj Close**, Volume) were extracted and converted into a CSV (comma-separated values) file.

Company (Ticker)	Year Added
3M* (NYSE:MMM)	1976
American Express (NYSE:AXP)	1982
Apple (NASDAQ:AAPL)	2015
Boeing (NYSE:BA)	1987
Caterpillar (NYSE:CAT)	1991
Chevron* (NYSE:CVX)	2008
Cisco (NASDAQ:CSCO)	2009
Coca-Cola* (NYSE:KO)	1987
The Walt Disney Company (NYSE:DIS)	1991
DowDuPont (NYSE:DD)	2017
ExxonMobil* (NYSE:XOM)	1928
General Electric (NYSE:GE)	1907
Goldman Sachs (NYSE:GS)	2013
The Home Depot (NYSE:HD)	1999
IBM (NYSE:IBM)	1979
Intel (NASDAQ:INTC)	1999
Johnson & Johnson* (NYSE:JNJ)	1997
JPMorgan Chase (NYSE:JPM)	1991
McDonald's* (NYSE:MCD)	1985
Merck (NYSE:MRK)	1979
Microsoft (NASDAQ:MSFT)	1999
Nike (NYSE:NKE)	2013
Pfizer (NYSE:PFE)	2004
Procter & Gamble* (NYSE:PG)	1932
Travelers Companies, Inc. (NYSE:TRV)	2009
United Technologies (NYSE:RTX)	1939
UnitedHealth (NYSE:UNH)	2012
Verizon (NYSE:VZ)	2004
Visa (NYSE:V)	2013
Wal-Mart (NYSE:WMT)	1997

Fig. 1. Rows including the 30 companies composed of the Dow Jones Industrial Average(DJIA). Column I consists of the company names with their tickers (in parentheses). Column II consists of the year that the companies were included in the index fund.

Next, for the data included in the CSV file, I calculate and extract the time series data of daily returns, R_d, for DJIA values.

$$R_d = (P_d - P_{d-1}) / (P_{d-1}) \qquad (1)$$

P_d represents the closing price of the stock at day *d*. Normal-returns are used to measure only the pure daily return values gathered from the index's closing price rather than log-returns as shown in (1).

However, since stock price data does not completely consist of information given daily due to the stock market being closed on weekends and public holidays, P_e, an approach for compensating for missing data was approximated. The P_e approximation was made using a simple technique. An average between days were taken to acknowledge the Weekend Effect, which describes the phenomenon in which stock returns on Mondays are often lower than the most recent preceding Friday [22].

$$P_e = (P_x + P_y) / 2 \qquad (2)$$

P_x represents the stock closing price of a certain day whether it be the day before the weekend or a holiday, and P_y , represents the stock closing price of the next value present. For the purpose of this study in evaluating the stock closing price for the weekend, the estimated value for the stock closing price on a Saturday and Sunday would be equal, and hence, the same for holidays (although, there were no closing stock market dates during the timeline of this study).

B. Extracting Twitter Data

The next source of data is from extracting relevant tweets fromTwitter, along with their sentiment (as discussed in appendix *C. Performing sentiment analysis*). Using a Twitter API library, the data was collected in a Python3 program as it is an effective program for extracting and applying sentiment analysis to tweets [16]. After applying and gaining developer access to the Twitter API, I was able to start extracting tweets. By importing and installing *Tweepy,* a Twitter API library, onto a Python program, I was able to obtain access to all public information posted online via Twitter. The specified data was collected in a month's timeline from March 18, 2020, to April 22, 2020. The timeline was chosen as the timeframe of the project as my limited access to the Twitter API only granted me current week's data for its Sandbox subscription (most basic and free subscription). Any other subscription to the Twitter API would have cost thousands, which, due to my limited resources, was not able to be obtained. A search query was utilized that consisted of multiple keywords that would filter out and only target certain tweets that would relate to DJIA and the companies composed of it.

Since only a portion of all tweets may pertain to the DJIA, keywords only relating to the companies and the DJIA itself were included when filtering and extracting the tweets.". Only the large, well-known companies and the keyword #DJIA were sought after. Because the Standard Search API, the Twitter API package I had administered the search under, only extends to a 500 character limit, therefore, preventing all the companies in the DJIA to be used as a search query (exceeding the character limit of 500). The well-known companies were established by first including all companies and deleting relatively smaller-cap companies when the character limit was exceeded.

In addition, there was a rest period after a search was administered. Each day of extracting tweets had a usage limit where 25,000 tweets, due to the package, were only able to be extracted.

C. Sentiment Classification

In Python3, multiple libraries were installed, including Tweepy (as stated in *Extracting Twitter Data).* The Natural Language Toolkit (NLTK), one of the packages installed onto Python3, consisted of means to interpret natural language processing (NLP). NLTK downloaded multiple libraries and programs that supported NLP in order to interpret subjective tweets in Twitter. NLTK offered libraries able for text processing libraries for classification in NLP [23].

TextBlob was an additional Python3 library for processing textual data. The Tweepy library along with its API were installed with TextBlob that enabled completing and using common natural language processing (NLP) tasks in performing sentiment analysis on tweets extracted using the Twitter API. TextBlob consists of an API of a lexicon-based model for identifying words and applying a sentiment value. Values of sentiment for each word in TextBlob were predetermined in the programming. Thus, TextBlob has many features such as, but not limited to sentiment analysis, classification using a Naive Bayes Decision Tree, Language translation and detection powered by Google Translate (although the tweets extracted were in English), and tokenization (converting texts into words and sentences) [24].

First, I had converted the tweets into a CSV file, i.e., parsing the input file to be able to utilize the file for a sentiment analysis model in TextBlob. The API library in Textblob was used to perform the natural language processes for sentiment analysis. The TextBlob library contained both a Naive Bayes analyzer that was used as an NLTK analyzer and a pattern analyzer that was dependent on libraries for patterns in tweets. Code in Python3 parsed the CSV file undergoing sentiment analysis. As the sentiment analysis model was programmed, the run of the program provided a result of the tweet in the form of a polarity and a subjectivity value.

When preprocessing tweets, URLs were removed as they tend not to represent any sentiment-relevant content but rather lead to sentiment content, not textually present in the tweet. In addition, cash-tags e.g., “$AAPL”, and user taggings e.g., “@example”, were removed to make a tweet related specifically to a stock in the DJIA the DJIA itself, and/or the user’s discussion on the stock, thus formatting a model for generalizing tweet sentiment independence. Letter repetitions used to mock traditional English language words were collapsed letter repetitions using TextBlob’s lexicon system to adapt words into real words using machine learning (e.g., “amazingggg” becomes “amazing”). Afterward, I followed a computation procedure by applying tokenization that derived texts in the tweet into a resulting lexicon value, and an *n*-gram construction, both offered in the TexBlob and NLTK library kits. I had not removed stop words, such as “not”,”and”,”the”, and “is” as the deletion could have altered the sentiment polarity of a tweet (which will be further discussed in V. Limitations/Conclusions).

Date	NumofTweets	NumofPositive	Numo Neutral	NumoNegative	PositiveMean	NegativeMean	OverallMean
3/18/2020	2487	823	1334	330	0.357	-0.267	0.178
3/19/2020	3132	1064	1696	372	0.343	-0.254	0.188
3/20/2020	4206	1234	2110	862	0.338	-0.182	0.124
3/21/2020	3118	976	1683	459	0.349	-0.221	0.167
3/22/2020	3083	1035	1667	381	0.366	-0.25	0.200
3/23/2020	3360	1131	1797	432	0.375	-0.262	0.199
3/24/2020	4033	1436	2110	487	0.37	-0.254	0.212
3/25/2020	3468	1165	1919	384	0.36	-0.286	0.200
3/26/2020	3547	1212	1911	424	0.368	-0.247	0.209
3/27/2020	5000	1674	2529	797	0.402	-0.173	0.217
3/28/2020	3038	968	1698	372	0.38	-0.284	0.196
3/29/2020	3070	985	1718	367	0.363	-0.284	0.187
3/30/2020	3591	1118	2067	406	0.361	-0.297	0.186
3/31/2020	3577	1170	1971	436	0.372	-0.27	0.198
4/1/2020	3471	1200	1883	388	0.385	-0.271	0.225
4/2/2020	3301	1082	1786	433	0.368	-0.26	0.189
4/3/2020	4111	1487	2123	501	0.36	-0.256	0.205
4/4/2020	3235	1068	1814	353	0.368	-0.285	0.206
4/5/2020	3202	1126	1734	342	0.389	-0.27	0.235
4/6/2020	3393	1032	1977	384	0.365	-0.282	0.190
4/7/2020	3642	1172	1955	515	0.367	-0.282	0.169
4/8/2020	3438	1094	1879	465	0.359	-0.239	0.181
4/9/2020	3537	1243	1880	414	0.349	-0.276	0.193
4/10/2020	2260	786	1203	271	0.367	-0.232	0.213
4/11/2020	1688	550	967	171	0.358	-0.282	0.206
4/12/2020	3221	1090	1764	367	0.404	-0.259	0.237
4/13/2020	3089	1032	1689	368	0.384	-0.258	0.215
4/14/2020	3325	1082	1866	377	0.362	-0.283	0.195
4/15/2020	3306	1080	1840	386	0.382	-0.261	0.213
4/16/2020	3349	1066	1916	367	0.364	-0.263	0.203
4/17/2020	3559	1264	1855	440	0.367	-0.251	0.207
4/18/2020	3202	1047	1784	371	0.377	-0.253	0.212
4/19/2020	3401	1013	2028	360	0.35	-0.277	0.186
4/20/2020	3611	1272	1957	382	0.37	-0.247	0.227
4/21/2020	3294	1034	1870	390	0.369	-0.279	0.191
4/22/2020	3905	1246	2219	440	0.378	-0.264	0.211

Fig. 2. Cells presenting resulting data from TextBlob and Tweepy. Rows represent the data associated with each date labeled in column 1. Column 2 represents the quantity of tweets extracted using the SandBox Standard Search Twitter API. Columns 3, 4, and 5 notes the number of tweets associated with a polarity sentiment value of positive, neutral, and negative, respectively. Columns 6, 7, and 8 represents the average/mean sentiment score for tweets considered positive, negative, and the overall mean sentiment score, respectively. (using MatPlotLib in Python3 to program the table).

TextBlob’s sentiment analysis structure was based on a Naive Bayes Classifier involving conditional probability. The theorem described a given precondition, and the corresponding probability of a score that was predetermined within the library. TextBlob considered the probability of words and phrases in a text being positive or negative based on recent information related to the text, thus, providing a lexicon-based method. For example, question marks typically meant a state of worry and wonder, thus, influencing the sentiment score of its relating text. After the summation of the values of sentiment for each text in the tweet, the outcome value was denoted as its score. The Naïve Bayes classifier consisted of the counting, multiplying, and calculating the probability between values which enabled a faster and more efficient runtime than other classifiers sorting tweets into sentiments. Despite an effective method, the accuracy of a Naïve Bayes classifier for sentiment analysis such as TexBlob was relatively similar, yet, distinctively undetermined to other text classifiers such as Google-Profile of Mood States (GPOMS) and Opinion Finder (OF) [11][25].

The results were stored via CSV file where quantities for each subjective and polarity value is stored, but the text of the tweets themselves are not. The sentiment values (polarity and subjectivity) helped determine the sentiments of the comment. Polarity values were calculated from -1.0 to 1.0 subjectivity values were calculated 0.0 to 1.0 (where the higher the value, the more subjective it was). However, subjectivity scores were disregarded in this study as the paper focuses on polarity and the significance of sentiment alone. The polarity score presented the value of the connected positive and negative information that was in the text of the tweet The average positive polarity of sentiment, the average negative polarity of sentiment, and the overall sentiment polarity were taken. These scores allowed me to generalize and understand the opinions of the vast majority of the public.

If polarity was greater than 0, then the text was considered positive. If the sentiment score was equivalent to 0, then the comment was considered neutral, thus, having no polarity of sentiment, and if the score was less than 0, then the tweet was considered negative.

The resulting dataset showed a display of 121,250 subsets of annotated tweets regarding the DJIA out of that were analyzed to capture 40,027 positive deemed tweets, 66,199 neutral tweets, and 15,024 negative tweets. Each average sentiment score was then estimated to 3 significant figures to stay consistent across values (further analysis was shown in IV. Results).

D. Granger Correlation Causality

In an analysis of the correlation between Twitter sentiment polarity values and DJIA stock prices, I applied a Granger causality test for the mean sentiment score.

Granger causality analysis is based on the assumption that if variable X causes variable Y, then the changes in X will have occurred before Y changes [11]. Thus, the basis for the assumption allowed me to determine significance by performing the Granger causality test. The changes in sentiment polarity would be determined to see if sentiment can be represented as a good model for predicting future stock prices in the DJIA.

$$\mathbb{P}|Y(t+1) \in A \mid I(t) \neq \mathbb{P}[Y(t+1) \in A \mid \mathrm{I}_{-X}(t)] \qquad (3)$$

Where $\mathbb{P}$ refers to probability, A is an arbitrary non-empty set, $I(t)$ and $\mathrm{I}_{-X}(t)$ respectively normalize information available as time, t [26]. For the Granger causality, (3) must be assumed to be true for X to cause Y [27].

Lagged values of X (sentiment change) founded by the Granger causality test will be tested to see if X exhibited a statistically significant correlation with Y changes in DJIA values illustrated in (1) with corresponding dates. Lag was determined by day(s), t. Lags included a length of days from 1-7, hence, provided a week lag for determining significance in both time series. Lag would be identified by t subtracted by 1-7 i.e., t-1, for a lag of 1 day. However, correlation does not necessarily translate to causation [28]. Thus, I performed the Granger causality test similar to [11] in a predictive fashion, rather than a causative fashion. The time series, thus, was tested, to determine if sentiment changes were predictive of changes in stock price, not causative of stock price change.

During the Granger causality test, a sum of squares (SSR) based F-test, SSR based chi-square test, likelihood ratio test, and a parameter F test were performed to determine potential differences in baseline fits and models (Granger causality test) to determine significance with a 99% confidence interval using an alpha value, α, of 0.01 .

IV. Results

An analysis of the Granger causality revealed that the correlation between Twitter sentiment polarity and the change in the DJIA stock return, R_d, was insignificant. By plotting/graphing corresponding values via Python3 and MatPlotLib (Python3 library for embedding plots), the graph is evident that while trends look similar, the correlation was small.

A. Granger Causality Test

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=0.2281  , p=0.6362  , df_denom=32, df_num=1
ssr based chi2 test:   chi2=0.2495  , p=0.6175  , df=1
likelihood ratio test: chi2=0.2486  , p=0.6181  , df=1
parameter F test:         F=0.2281  , p=0.6362  , df_denom=32, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=0.0367  , p=0.9640  , df_denom=29, df_num=2
ssr based chi2 test:   chi2=0.0861  , p=0.9579  , df=2
likelihood ratio test: chi2=0.0860  , p=0.9579  , df=2
parameter F test:         F=0.0367  , p=0.9640  , df_denom=29, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=0.0760  , p=0.9724  , df_denom=26, df_num=3
ssr based chi2 test:   chi2=0.2892  , p=0.9620  , df=3
likelihood ratio test: chi2=0.2880  , p=0.9623  , df=3
parameter F test:         F=0.0760  , p=0.9724  , df_denom=26, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test:         F=1.8159  , p=0.1601  , df_denom=23, df_num=4
ssr based chi2 test:   chi2=10.1057 , p=0.0387  , df=4
likelihood ratio test: chi2=8.7823  , p=0.0668  , df=4
parameter F test:         F=1.8159  , p=0.1601  , df_denom=23, df_num=4

Granger Causality
number of lags (no zero) 5
ssr based F test:         F=1.7202  , p=0.1760  , df_denom=20, df_num=5
ssr based chi2 test:   chi2=13.3315 , p=0.0205  , df=5
likelihood ratio test: chi2=11.0889 , p=0.0496  , df=5
parameter F test:         F=1.7202  , p=0.1760  , df_denom=20, df_num=5

Granger Causality
number of lags (no zero) 6
ssr based F test:         F=1.2843  , p=0.3160  , df_denom=17, df_num=6
ssr based chi2 test:   chi2=13.5984 , p=0.0345  , df=6
likelihood ratio test: chi2=11.2147 , p=0.0820  , df=6
parameter F test:         F=1.2843  , p=0.3160  , df_denom=17, df_num=6

Granger Causality
number of lags (no zero) 7
ssr based F test:         F=0.3253  , p=0.9297  , df_denom=14, df_num=7
ssr based chi2 test:   chi2=4.7166  , p=0.6945  , df=7
likelihood ratio test: chi2=4.3701  , p=0.7363  , df=7
parameter F test:         F=0.3253  , p=0.9297  , df_denom=14, df_num=7
```

Fig. 3. List of Granger causality test results. Figure includes statistical tests: SSR based F- test, SSR based chi2 test, likelihood ratio test, and parameter F-test. Values given from the test are F-values, p-values, chi2-values, etc. (as shown above).

Though, with a 99% confidence interval, looking towards the statistical tests in Fig. 3., all F-values, chi2, and p-values are greater than 0.01; thus, deeming the data insignificant.

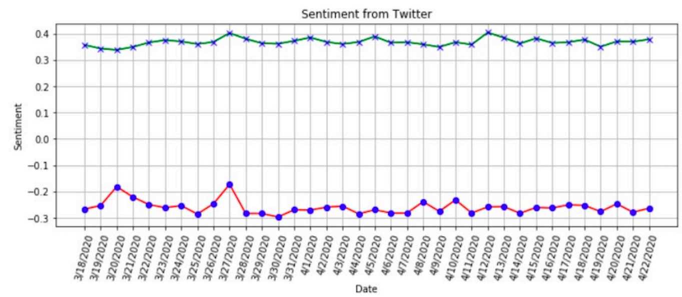
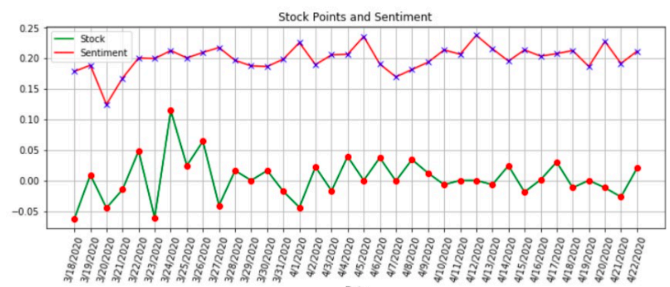


Fig. 4. Graph representing average sentiment of both positive and negative polarity tweets. Green line with blue x-marks is the average polarity of positive sentiment and the red line with blue dots is the average polarity of negative sentiment.

Fig. 5. Normalization of Daily stock returns in DJIA (green line with red dots) and average polarity sentiment (positive and negative in red line with blue x-marks).



With an overall and abundant average polarity sentiment score as shown in Fig. 4, Fig. 5. showed slight to no resemblance in pattern between stock returns in DJIA and sentiment change.

However, Fig. 3 showed that a lag of 2 days (t-2) would have represented the best model for a prediction model for DJIA stock returns as its statistical values were closest to the confidence interval of 99%.

Therefore, the polarity sentiment variable was not useful for predicting the price return, as the statistical values yielded an insignificant result. Though, the number of tweets for a company could have not been considered as inaccurate for a lag of 2 days if α was increased to a significance level of 0.05. Thus, Twitter is proved useful for predicting volatility in price for stocks when more attention is drawn to tweets.

B. Stock Prediction LSTM Non-Linear Model

Because the Granger causality test revealed that sentiment change was not predictive of stock changes in the DJIA, a machine-learning technique for predicting stock prices was used of purely stock data of previous historical DJIA values. [11] had used a Soft Organizing Fuzzy Neural Network (SOFNN) to predict future stock movements, however, a recurrent neural network (RNN) has not, despite both being effective machine learning methods for non-linear methods. Since changes in Twitter sentiment was not predictive, the limited timeline from the Twitter API was not necessary. A greater timeline would yield more accurate results using a Long Short Term Memory (LSTM) model for more accurate prediction movements. An LSTM model follows a deep learning framework that remembers time intervals over an extended period of time, and is an effective model for plotting and measuring time-series data [29]. A timeline of March 12, 2016 to April 29, 2020, was chosen as the timeline to effectively cover at least 4 years of historical data in DJIA values for a more accurate prediction model and dataset.

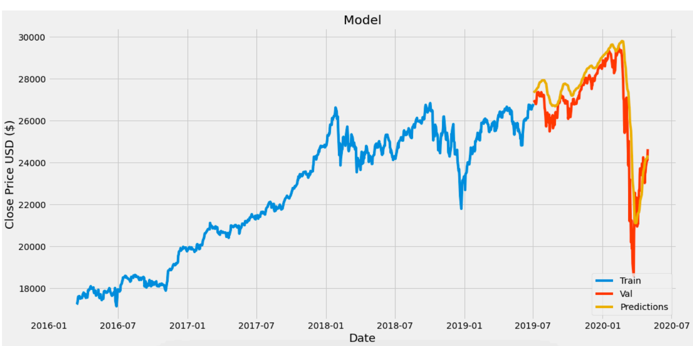


Fig. 6. Daily Dow Jones Industrial Average values between March 12, 2016 and April 29, 2020. Values (in red and blue), training data (in blue), and prediction graph (in yellow).

TABLE I

REAL AND ESTIMATE PREDICTION VALUES OF DJIA MARCH 12, 2016 TO APRIL 29, 2020

Date	Close	Predictions
2019-07-03	26966.000000	27326.835938
2019-07-05	26922.119141	27370.783203
2019-07-08	26806.140625	27415.021484
2019-07-09	26783.490234	27448.781250
2019-07-10	26860.199219	27473.119141
...	...	...
2020-04-23	23515.259766	24105.498047
2020-04-24	23775.269531	24149.416016
2020-04-27	24133.779297	24194.691406
2020-04-28	24101.550781	24262.236328
2020-04-29	24633.859375	24336.937500

208 rows × 2 columns

I used a 60-day training window for future prediction movements as shown in Fig. 6. where yellow represents the prediction model. Then, the LSTM model was built to have two LSTM layers with 50 neurons. Two Dense layers were also built with one layer containing 25 neurons and the other with 1 neuron. After training the data set and retrieving a prediction model as shown in Fig. 6., a Root Mean Square Error (RMSE) was calculated to determine the accuracy of the estimated prediction value with the real closing price value of the DJIA.

From the given dataset, I received an RMSE value of 1227.6668584872302, thus, showing a slightly accurate model of prediction. While quantitatively, an accurate quote of a DJIA value was inaccurate, a relatively close and visible relationship can be observed in Fig. 6. And Table I in terms of general trends and direction.

V. Discussion

A. Conclusion

In this paper, I investigated whether Twitter sentiment regarding the Dow Jones Industrial Average index fund and the companies composed of DJIA was predictive of stock returns specific to DJIA. Based on that result, I aimed to build a prediction model using an LSTM model using either previous historical data of DJIA or a normalization of data between Twitter sentiment and DJIA values. With an F-statistic, chi2-value, and p-value > 0.01, the data is insignificant, thus, Twitter feed and the sentiment within it has a slight to no correlation with changes in stock return in the DJIA. However, using an LSTM model for further predictions of DJIA, there was a directional accuracy of an RMSE value of 1227.6668584872302, yielding observable trends, yet quantitative accuracy. Thus, this research leads to two primary conclusions: 1) the use and development of machine-learning techniques for under-utilized data such as tweets are to be considered useful for us in terms of the economy and other items that apply to the country, and 2) a dismissal of the EMH in order to progress understanding of economics and societal factors that change the stock market.

B. Limitations

Despite a rejection of the hypothesis, there were several limitations to my study that are evident. First, was the lack of efficient and more accurate materials. Besides a Sandbox Standard Search API offered by Twitter developers, there were several other packages that were able to be used for more accurate data and more tweets to be analyzed. However, due to a premium subscription of $2499/month, a premium subscription for access to the Twitter API in Tweepy was inaccessible.

Another limitation to my project was the timeframe that related to the accessibility of the subscription package in Tweepy. Because the Sandbox package only allows prior time of 1 week when extracting tweets, only a few months of limited tweets were collected. With Twitter's premium subscription for API usage, I would have access to the entire Twitter database, rather than 1 week's worth. In addition, a premium subscription package to the Twitter API would increase my query length from 256 characters to 1024 characters, which would have permitted me to extract tweets regarding all companies associated with DJIA rather than a selected few. For an upgraded package, an increased query length and time frame would have allowed me to extract more tweets and more accurate tweets to calculate a more accurate model of average sentiment scores to determine public mood.

TextBlob's library was also a potential limitation. As the library was lexicon-based and the applied sentiment to each word was predetermined, the accuracy of sentiment applied to each tweet may have been inaccurate, which would have skewed whether the data compared to DJIA values were significant.

Thus, taking into account a premium API subscription package and more sentiment analysis techniques may have yielded more accurate results.

C. Implications

Because the statistical values during the Granger causality test were insignificant compared to an alpha value that is less than 0.01, sentiment was not predictive. However, there is the implication that if the confidence interval was decreased to a 95% interval for α to be set to 0.05, the data would have been significant enough to be predictive of stock return in DJIA. While [11] found an 86.7% directional accuracy between sentiment and DJIA values, despite having used multiple different sentiment analysis techniques, the claim can be made that future research in which changing variables in terms of different sentiment analysis techniques, correlation tests of time series, and specific Twitter data could improve the accuracy of stock market predictions.

D. Future Studies

More research regarding technical analysis in the stock market should be conducted when evaluating investment strategies for stock market predictions. With a smaller dataset and query length due to the limited API package offered from Twitter, I plan to use the premium package in the future for a more accurate dataset depicting tweets towards DJIA, giving me more tweets, and a better model for the representation of public mood in the form of sentiment polarity.

In addition, since the subjectivity score was not looked into during this study, future research could suggest whether subjectivity values may have affected and correlated with an insignificant model for predicting stock trends.

Future research regarding other sentiment analysis techniques may also be done to determine the accuracy and effectiveness of TextBlob compared to other methods which may have altered the average polarity score shown in Fig. 4.

Lastly, while DJIA presents an effective representation of the diversification of the stock market, individual stocks and tweets associated with the companies may be looked into to determine whether tweets regarding an individual company is predictive of the company's stock price. By spreading the process through multiple different technique approaches, technical analysis would be closer to determining whether social media and public mood are predictive of future stock market movements.

REFERENCES

[1] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology 2009.

[2] Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014). Twitter and society (Vol. 89). Peter Lang.

[3] Fama, E., Fisher, L., Jensen, M., & Roll, R. (1969). The Adjustment of Stock Prices to New Information. International Economic Review, 10(1), 1-21. doi:10.2307/2525569

[4] Fama, E. (1965). The Behavior of Stock-Market Prices. The Journal of Business, 38(1), 34-105. Retrieved from http://www.jstor.org/stable/2350752

[5] Malkiel, B. G. (2003). The efficient market hypothesis and its critics. Journal of economic perspectives, 17(1), 59-82.

[6] Qian, Bo, Rasheed, & Khaled. (2007) Applied Intelligence 26, 25–33.

[7] Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005, August). The predictive power of online chatter. In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (pp. 78-87).
ACM.

[8] Sadi, R., Asl, H. G., Rostami, M. R., Gholipour, A., & Gholipour, F. (2011). Behavioral finance: the explanation of investors' personality and perceptual biases effects on financial decisions. International journal of economics and finance, 3(5), 234-241.

[9] Dolan, R. J. (2002) Science 298, 1191–1194.

[10] Nofsinger, J. (2005) Journal of Behaviour Finance. 6, 144–160.

[11] Bollen, Johan & Mao, Huina & Zeng, Xiao-Jun. (2010). Twitter Mood Predicts the Stock Market. Journal of Computational Science. 2. 10.1016/j.jocs.2010.12.007.

[12] Gallup, G. (Ed.). (1994). The gallup poll: public opinion 1993. Rowman & Littlefield.

[13] Edmans, A., Garcia, D., & Norli, Ø. (2007). Sports sentiment and stock returns. The Journal of Finance, 62(4), 1967-1998.

[14] Hirshleifer, D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. The Journal of Finance, 58(3), 1009-1032.

[15] Gilbert, E., & Karahalios, K. (2010, May). Widespread worry and the stock market. In Fourth International AAAI Conference on Weblogs and Social Media.

[16] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).

[17] Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. Journal of happiness studies, 11(4), 441-456.

[18] Butler, K. C & Malaikah, S. J. (1992) Journal of Banking and

Finance 16, 197–210.

[19] Kavussanos, M & Dockery, E. (2001) Applied Financial Economics 11, 573–79.

[20] Pak, A & Paroubek, P. (2010) Twitter as a Corpus for Sentiment Analysis and Opinion Mining. (European Language Resources Association (ELRA), Valletta, Malta).

[21] Dow Jones Industrial Average (^DJI) Charts, Data & News. (2020, April 19). Retrieved April 22, 2020, from https://finance.yahoo.com/quote/^DJI/

[22] Keim, D., & Stambaugh, R. (1984). A Further Investigation of the Weekend Effect in Stock Returns. The Journal of Finance, 39(3), 819-835. doi:10.2307/2327945

[23] Almohaimeed, A. (2017, August 5). Using Tweets Sentiment Analysis to Predict Stock Market Movement. Retrieved January 1, 2020, from https://etd.auburn.edu/bitstream/handle/10415/5759/Using Tweets Sentiment Analysis to Predict Stock Market Movement.pdf?sequence=2

[24] Kapadia, F. (2017, May). Sentiment Analysis: Android applications. Retrieved January 2020, from https://www.cs.rit.edu/usr/local/pub/GraduateProjects/2165/f jk9481/Report.pdf

[25] Sarkis Agaian, Petter Kolm. Financial Sentiment Analysis Using Machine Learning Techniques. International Journal of Investment Management and Financial Innovations. Vol. 3, No. 1, 2017, pp. 1-9.

[26] Busru, S. A., & Shanmugasundaram, G. (2017). Effects of Innovation Investment on Profitability and Moderating Role of Corporate Governance: Empirical Study of Indian Listed Firms. Indian Journal of Corporate Governance, 10(2), 97–117. https://doi.org/10.1177/0974686217730938

[27] Eichler, Michael (2012). "Causal Inference in Time Series Analysis" (PDF). In Berzuini, Carlo (ed.). Causality : statistical perspectives and applications (3rd ed.). Hoboken, N.J.: Wiley. pp. 327–352. ISBN 978-0470665565.

[28] Glymour, M. M. (2006). Using causal diagrams to understand common problems in social epidemiology. Methods in social epidemiology, 393-428.

[29] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735