

# Does Twitter affect Stock Market Decisions? Financial Sentiment Analysis in Pandemic Seasons: A Comparative Study of H1N1 and COVID-19

David Valle-Cruz (✉ [davacr@uaemex.mx](mailto:davacr@uaemex.mx))

Universidad Autónoma del Estado de México <https://orcid.org/0000-0002-5204-8095>

Vanessa Fernandez-Cortez

Universidad Autónoma del Estado de México <https://orcid.org/0000-0002-4572-1544>

Asdrúbal López-Chau

Universidad Autónoma del Estado de México <https://orcid.org/0000-0001-5254-0939>

Rodrigo Sandoval-Almazan

Universidad Autónoma del Estado de México <https://orcid.org/0000-0002-7864-6464>

---

## Research Article

**Keywords:** Sentiment Analysis, Affective Computing, Finance, Stock Market, Pandemic, Twitter

**Posted Date:** July 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-39991/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Cognitive Computation on July 7th, 2020.  
See the published version at <https://doi.org/10.1007/s12559-021-09819-8>.

# Abstract

**Background:** Investors are always playing with the fears and desires of buyers and sellers. Stock exchange markets are not the exception. Financial sentiment analysis allows us to understand the effect of reactions and emotions on social media in the stock market. In this research, we analyze Twitter data and financial indices to answer the question: How do polarity generated by the posts on Twitter influence financial indices behavior in pandemic seasons?

**Methods:** The study is based on the sentiment analysis of influential Twitter accounts in this field and its relationship with the behavior of important financial indices. To achieve this, we tested four lexicons to detect polarity on Twitter.

**Results:** Our findings shows that the period in which the markets reacted was 6 to 13 days after the information was shared and disseminated on Twitter in the COVID-19 season, and 1 to 2 day for H1N1 season. Furthermore, in our analysis, we found that the lexicons that got the best results for sentiment analysis on Twitter were S140 and Affin.

**Conclusions:** Financial sentiment analysis is an important technique to forecasting stock market and polarity is the most widely used technique in the financial area. There is a relationship between the polarity in Twitter and the financial indexes behavior. The most influential Twitter accounts during the pandemic season were The New York Times, Bloomberg, CNN News, and Investing, presenting a very high relation between sentiments on Twitter and the stock market behavior.

## Introduction

Financial markets are based on expectations. As stated earlier, investors are always playing with the fears and desires of buyers and sellers. Kiyamaz [17] studied 355 favorable tweets; agreeing that rumors related to earning expectations, and purchases by foreign investors generated a more significant impact on stock prices. Cruz and Gómez [9] confirmed Kiyamaz [17] opinion and added that investors make decisions based on rumors betting on the credibility of the media that disclose them, even though knowing that the information is not always reliable. However, this trend grows with the use of online social media networks such as Twitter, Facebook, Instagram, and YouTube, which increases the exposure of topics and ideas through their platforms impacting the stock market. Emergencies are special cases, where the amount of available information is reduced, and decisions have to be made on such bases. Recently, pandemic COVID-19 changed our lives and altered the balance among stock markets around the world and created an unstable environment for several weeks. The same thing happened 11 years ago, with the pandemic caused by another the virus H1N1.

Emotions and sentiments are inherent in investors' reactions and decision-making, learning, communication, and awareness [26]. Understanding emotions is one of the most important aspects of personal development and growth, and their understanding is paramount to the emulation of human intelligence. Thus, the processing of emotions is important for the advancement of artificial intelligence

and is a task closely related to the detection of polarity and emotions [6]. One of the most widely used techniques for sentiment analysis is machine learning, which has allowed us to analyze a large amount of data. In this regard, data based on video, image, audio, and text are analyzed in the BigData, complementing its analysis with relational databases. The potential of this type of technique has extended to different fields, such as social media, where sentiment analysis or opinion mining is applied to finding patterns of behavior, identifying emotions, trends, and even inducing people's decisions.

Recently, research on sentiment analysis on finance has increased, but very few studies use social media datasets. However, it is important to identify signs or patterns that guide investors when selecting the assets that will make up their portfolios, trying to ensure that the prices of the selected stocks have a very similar trend or above the prices of the corresponding indices. Therefore, the investor must have tools that allow him/her to consider the most considerable information available for a technical and fundamental analysis that leads him/her to reduce risk and uncertainty as much as possible, as well as a maximum return.

The shared information in social media is a source that generates sentiments and reactions in investors, which influence their decisions to buy or sell financial assets and especially in the stock market [16]. News shared on social media, whether true or not, causes changes in the trends of international stock market indices. Some studies confirm the relationship between sentiments generated by social media in investors and market trends, and this research continues this path in pandemic seasons.

The purpose of this investigation is to compare these emergency events with similar characteristics: a health situation, epidemic distribution, and worldwide attention, and the use of social media platforms that improve their diffusion of rumors and news that alter investors' perceptions on their stock exchange. The question guiding the document is: How do polarity generated by the posts on Twitter influence financial indices behavior in pandemic seasons? To answer this question, our analysis proposes a methodology based on the sentiment analysis of influential Twitter accounts in the financial field and its relationship with the behavior of important financial indices. We analyzed two crucial moments in the global financial crises: the H1N1 pandemic of 2009 and the COVID-19 pandemic in 2020.

The document is divided into five sections: in the first section, we present a literature review based on affective computing and sentiment analysis in finance. In the second section, our methodology based on the analysis of the evolution of the selected financial indices and the sentiment analysis Twitter. The third section describes the findings. The fourth section is the discussion of results. Finally, we present the conclusions of the study and future work.

## Literature Review

The literature review consists of three sections. First, the literature on emotional computing and sentiment analysis; second section, some works related to the area of affective computing and finance, and the final section describes research on sentiment analysis and indices.

## 1. Affective Computing and Sentiment Analysis

Emotions are fundamental for a successful and effective communication between human beings. In fact, in many everyday situations, emotional intelligence is more important than intelligence quotient (IQ) for a successful interaction [4]. A modern and massive source for finding people's sentiments is social media; for instance, the content published on Facebook and Twitter shows the emotions and sentiments of people depending on any kind of event.

Affective computing and sentiment analysis are multidisciplinary areas that include academics and professionals from different disciplines, such as computer science, cognitive science, political science, economics, and finance [5, 26]. Particularly, sentiment analysis is an area of artificial intelligence, which allows us to extract and analyze data stored in social media, images, sounds, and videos. The sentiment analysis allows finding patterns or characteristics in the information published in long data sets, which can be very useful for decision-making in organizations, political movements, business strategies, marketing campaigns, and product preferences, among others [26, 27]. The sentiment analysis reveals personal opinions towards entities such as products, services, or events, which can affect organizations and companies by improving their marketing, communication, production, and acquisition [18].

Cambria et al. [6] argued that sentiment analysis had been confused with the task of polarity detection: from negative to positive sentiments. However, this is only one of many natural language processing problems that must be solved to achieve human-like performance in sentiment analysis. Thus, when performing sentiment analysis, we encounter a complex problem of natural language processing that can be analyzed from different emotional or psychological models [5, 23].

The opportunity to automatically capture the sentiments of the public about social events, political movements, marketing campaigns, and product preferences has raised a growing interest both in the scientific community, for the exciting open challenges, and in the business world, for the remarkable repercussions on marketing and financial market prediction sentiments [7].

This has led to the emergence of the fields of affective computing and sentiment analysis, which take advantage of human-computer interaction, information retrieval, and multimodal signal processing to distill people's emotions from the growing amount of social data online.

## 2. Sentiment Analysis and Finance

As argued in the previous section, sentiment analysis is the computational study of people's opinions, emotions, assessments, attitudes, moods, and emotions. It has become one of the most active areas of research in natural language processing, data mining, information retrieval, and Web mining [21]. In recent years, its research and applications have also extended to the managerial and social sciences, because of their importance for businesses, governments, and society.

Affective computing and sentiment analysis have great potential to improve the capacity of customer relationship management and referral systems. For example, to reveal what characteristics clients enjoy,

or to exclude items from referrals that were received with a negative response. Business and financial intelligence is also a major factor in the interest of companies in the elements of affective computing and sentiment analysis [4].

The rapid progress of the Internet has had a significant impact on the financial field. How to quickly and accurately extract key information from negative financial texts has become one of the key issues for investors and decision-makers [34]. Investors have always been interested in stock price forecasting; however, little research has been done on social media and its effects on forecast results in pandemics seasons. In this regard, Day and Lee [10] argued that financial sentiment analysis is an important area of financial technology research (FinTech). Chen and colleagues [8] found that social media sentiment associates strongly with stock returns and identified the importance of social media as an additional channel to reflect stock prices.

Some studies analyze Twitter data for the sentiment analysis applied to finance. Souza et al. [31] investigated the relationship between sentiment and volume of data on Twitter, and the profitability and volatility of the stock market, finding that social media are a valuable source in analyzing the financial dynamics of the retail sector, even when compared to major news outlets such as The Wall Street Journal and the Dow Jones Newswires. Ahmed [2] argued that investors' reaction to a financial event, in the form of stock price movement, reflects the severity of the event and that the words used in the published news show a similar degree of emotion.

Gomez, Prado, and Plaza [14] generated an algorithm based on artificial intelligence that uses investors' sentiment to open long or short positions in the future Ibex 35. To measure investors' sentiment, the authors used Semantic Analysis Algorithms that qualified as good, bad, or neutral any communication related to the Ibex 35 made on Twitter or in the media [28]. The authors demonstrated that by assessing investor sentiment, they could predict the evolution of the Ibex 35 and improved their returns relative to Sharpe, success rate, and a profit factor of traditional trading systems.

In the financial sector, the application of sentiment analysis has focused on identifying the polarity related to the similarity or relationship between the upward - downward financial trends and the sentimental polarity that varies from positive to negative. Thus, Devitt and Ahmad [12] analyzed how financial news texts affect investors' judgment, which impacts transaction volumes, stock prices, volatility, and even future company earnings, by classifying the information obtained into positive and negative. De la Orden, Martínez, and Vianez [11] found differences in the sentiments transmitted by the Spanish media, in digital format, specialized and general. They also observed that the media that transmitted less sentiment is mainly negative.

The data generated in social media have become a valuable resource for the analysis of sentiment in the financial field, as they have proved to be extremely important for market research companies and public opinion organizations. In this regard, Atzeni, Dridi, and Recupero [3, 13] suggested a continuous polarity scoring approach in the range of  $[-1, +1]$  to identify bullish and bearish sentiment associated with companies and stocks, finding that the ontological framework-based resource could be successfully

applied for sentiment analysis within the financial domain. They also achieved better results than traditional methods of sentiment analysis that do not incorporate semantics.

Smailović et al. [29] used predictive sentiment analysis to forecast closing stock price movements via Twitter feeds, finding that the polarity of sentiment can indicate stock price movements a few days in advance.

According to Zhang et al. [33], the daily happiness generated on Twitter influences the performance and volatility of international indices. In this paper, they examined the relationships between the everyday feeling of happiness on Twitter and stock market performance in 11 international stock markets, dividing this feeling of happiness into quintiles from the least to the happiest days. They found that the correlation coefficients between the feeling of happiness and the index performance in the happiest subgroups 4 and 3 were higher than in the last 2 and 3 happiness subgroups and that the sentiment of happiness could provide additional explanatory power for the return index in the happiest subgroup. Third, daily happiness can increase changes in the index performance for most stock markets. Fourth, we found that the return of the index and the volatility based on the range of the happiest subgroup are higher than those of other subgroups.

### 3. Sentiment Analysis and Social Media in the Stock Market

Sentiment analysis and its effect on the stock market have produced many studies by scholars around the world [15]. Kearny and Liu [16] argued that future research on finance is related to the availability of more accurate and efficient sentiment measures and more wide-ranging studies. Chen, et al. [8] found that social-media sentiment is strongly associated with stock returns and the importance of social media as an additional channel where stock prices are reflected.

Ruiz-Martinez and colleagues [28], pioneer of studies on sentiment analysis over the financial market, proposed an algorithm for opinion extraction in financial news developing an ontology to understand this object of study. Another contribution to the field found that the financial community behaves similarly to a small-world network and identified critical nodes to analyze its influence [35]. They used a novel sentiment analysis algorithm measuring tweet messages from the nodes and discover that it is significantly correlated to the returns of the major financial market indices. Souza and colleagues [31] also supported the idea that social media is valuable for the analysis of financial dynamics, using the retail industry analyzing twitter sentiment and volume, and stock returns and volatility.

Research from Liew et al. [19] trained several standard machine learning classification algorithms, and found that support vector machines (SVM) generate the most useful predictive sentiments. They also found a strong positive relation with tweet sentiments on daily market returns, and a weaker negative correlation with next day market returns. Another contribution from Liew and Budavari [20] is that they identified a social media factor showing that tweet sentiments have significant power in explaining the time-series contemporaneous variation in daily stock returns.

The most recent research comes from Schangir et al. [30] which uses deep learning to analyze big data in StockTwits using short-term memory, doc2vec, and convolutional neural networks [25]. Their findings support the argument that deep learning and convolutional neural networks are the best models to predict the sentiment of authors in StockTwits data sets. The above-mentioned studies frame our research object of study, and our argument to analyze financial markets using different tools of artificial intelligence.

Finally, similar studies to our research sample that uses indices data such as Standard & Poor's (S&P) started with Sul and colleagues [36]. They introduced a research from the Standard & Poor's 500 index collecting twitter posts from the firms on this index and compared them with the average daily stock market returns. They found that the cumulative emotional valence of tweets about a specific firm was significantly related to that firm's stock returns. Another finding was in terms of the number of followers. If the emotional level of tweets with many followers (more than the median) had a stronger impact on same day returns, then the emotional level of tweets with few followers had a stronger impact on future stock returns (10-day returns).

Sun and Fabozzi [32] tested their model using the majority of stocks listed on the S&P 500 index. They did not evaluate sentiment, but they analyzed textual information from microblogs. They correlated movements of both stock prices and social media content. All of this is evidence that there are different perspectives in the study of financial sentiment analysis to understand two main domains: social media and finance.

## Methodology

This section describes the methodology of the study based on the comparison of different lexicons to determine the polarity on Twitter and its relation to the financial indices behavior, and it is divided into two stages. First, the analyzed data of the stock market and the selected Twitter accounts. Secondly, the process of extracting and analyzing data to determine polarity.

### 1. Description of Data

The methodology is based on the sentiment analysis applied to Twitter accounts that influence the stock market behavior. To achieve this, tweets were downloaded at two important moments: H1N1 pandemic and COVID-19 pandemic. For the H1N1 season, the period considered was from June to July 2009 and for COVID-19 from January to May 2020. The above periods were selected from the time when the first stock market index peaked and began to fall. For instance, the point at which the pandemics started to influence the indices causing prices to fall to a minimum price, which marked the lowest prices reached as a result of the pandemic; and the period of financial disruption and uncertainty. Because the COVID-19 pandemic has been most damaging to the financial market, we extended the Twitter data analysis period to May 2020.

The financial indices were selected considering that they were the most representative and influential of each continent, in addition to including companies from different sectors to have a general perception of

their performance and behavior.

The considered financial data consisted of adjusting closing prices, where each of the financial indices in the world reached their maximum and minimum, first in the H1N1 influenza period and then in the COVID-19 period. The analyzed stock market indices are IPC, S&P 500, NASDAQ 100, Dow Jones, FTSE 100, Bovespa, CAC 40, DAX, Hang Seng, Nikkei 225 and SSE Composite (Table 1).

Table 1. Financial indicators description

Country	Index	Stock Market	Involved Sectors
Mexico	IPC	Bolsa Mexicana de Valores	Telecommunications, Consumer Products, Materials, Financial Services, Industrial, Healthcare, Financial Services
United States of America	NASDAQ 100	NASDAQ	Telecommunications, Biotechnology, Video Games, Hardware, Software, Internet Services, Financial Services, Pharmaceuticals, E-Commerce, Data Analysis
United States of America	Dow Jones	New York Stock Exchange (NYSE) Euronext	Covers all sectors except transport and utilities
United States of America	S&P 500	New York Stock Exchange (NYSE) Euronext	Consumer Discretionary, Health Care, Industrials, Information Technology,

			Materials, Real State, Communication
			Services, Utilities, Financials, Energy
United Kingdom	FTSE 100	London stock Exchange	Financial Services, Mining, Food, Pharmacy and Technology, Software and Computer Services. Life Insurance, Services
Brazil	Bovespa	Bolsa de Valores de Sao Paulo	Energy, Logistics, Finance and Insurance, Real Estate, Energy, Steel and Metallurgy, Aviation, Telecommunications, Food
France	CAC 40	NYSE Euronext	Energy, Logistics, Banking, Real Estate, Energy, Oil & Gas, Aerospace, Telecommunications, Food, Oil Equipment, Industrial Machinery
Germany	DAX	Frankfort	Logistics, Finance & Insurance, Real Estate, Energy, Pharmaceutical Chemicals, Airlines, Telecommunications, Consumer Goods, Automotive, Utilities, Building Materials, Technology, Software
Japan	Nikkei 225	Tokio Stock Exchange	Technology, Consumer Goods, Food, Transportation, Telecommunications, Transportation, Automotive, Tourism, Water, Construction, Healthcare, Finance, Electrical Industry
China	Hang Seng	Hong Kong Stock Exchange	Commercial Services, Communications, Consumer Goods, Technology, Energy Minerals, Transportation, Utilities, Health Services
China	SSE Composite	Shanghai Stock Exchange	Communications, Bank, Energy, Technology, Industrial, Finance & Insurance, Petrochemical, Food

Once the dates of the maximum and minimum prices were identified in the stock market behavior, the prices, from January 2009 to May 2020, were downloaded from the Yahoo Finance site and [Investing.com](https://www.investing.com), and the percentage of loss of each of the indices was calculated (Table 2). The adjusting closing prices were selected as they represent the closing price and corporate earnings to which you are entitled once you acquire the share. The adjusting closing price was used when analyzing historical returns and was calculated by the stock exchange to which it belongs.

Table 2. Maximum and minimum prices of the stock exchange indices

During the H1N1 season					
Index	Maximum price	Maximum price	Minimum price	Minimum price	Loss
	date		date		percentage
IPC	11/06/09	25372.84961	07/07/09	23359.9395	-7.9333232
NASDAQ 100	11/06/09	1862.369995	07/07/09	1746.17004	-6.2393591
Dow Jones	11/06/09	8770.919922	10/07/09	8146.52002	-7.1189785
S&P 500	11/06/09	944.890015	10/07/09	879.130005	-6.9595412
FTSE 100	10/06/09	4,436.80	10/07/09	4,127.20	-6.9780022
Bovespa	12/06/09	53558	14/07/09	48873	-8.747526
CAC 40	11/06/09	3334.939941	10/07/09	2983.1001	-10.55011
DAX	11/06/09	5107.259766	10/07/09	4576.31006	-10.39598
Hang Seng	11/06/09	18791.0293	13/07/09	17254.6309	-8.1762335
Nikkei 225	12/06/09	10135.82031	13/07/09	9050.33008	-10.709446
SSE Composite	10/06/09	2,816.25	12/06/09	2,743.76	-2.5739902
During the COVID-19 season					
Index	Maximum price	Maximum price	Minimum price	Minimum price	Loss
	date		date		percentage
IPC	12/02/20	45338.37109	23/03/20	32964.2188	-27.2928913
NASDAQ 100	19/02/20	9817.179688	20/03/20	6879.52002	-29.9236620
Dow Jones	12/02/20	29551.41992	23/03/20	18591.9297	-37.0861714
S&P 500	19/02/20	3386.149902	23/03/20	2237.3999	-33.9249600

FTSE 100	19/02/20	7,457.02	20/03/20	5,190.78	-30.3906923
Bovespa	19/03/20	68332	20/03/20	67069	-1.8483287
CAC 40	19/02/20	6111.240234	18/03/20	3754.84009	-38.5584604
DAX	19/02/20	13789	18/03/20	8441.70996	-38.7793897
Hang Seng	17/02/20	27959.59961	23/03/20	21696.1309	-22.4018542
Nikkei 225	12/02/20	23861.21094	19/03/20	16552.8301	-30.6287090
SSE Composite	13/01/20	3,115.57	23/03/20	2,660.17	-14.6169080

To select the Twitter accounts, it was considered that they would include special publications in the area of financial markets, finance, and economy. Additionally, the media where experts in the field commented, as well as places on the Internet that published general and international news that affected the world's population and therefore affected the sentiment of investors. In this manner, Twitter accounts selected for the sentiment analysis consisted of accounts of individuals, companies or organizations, and news broadcasts that were influential or important for finance. Some of the financial influencers do not have a Twitter account, so we downloaded Tweets from the company or organization they run, if there was one; furthermore, some Twitter users did not have an account in 2009 (Table 3).

Table 3. Analyzed Twitter accounts

Twitter account	Total downloaded tweets	
	2009	2020
@business	50	551
@Carl_C_Icahn	NA*	2
@CNNBUSINESS	747	1725
@ecb	NA*	298
@Investingcom	NA*	279
@InvestOfficeAD	NA*	120
@jpmorgan	NA*	86
@JPMorganAM	NA*	30
@lloydblankfein	NA*	8

@nytimes	490	244
@SEC_Enforcement	NA*	2
@UBS_CEO	NA*	9
@USTreasury	NA*	78
@WarrenBuffett	NA*	0

\*Not applicable, because the user did not have a Twitter account

## 2. Data Extraction and Analysis

Tweets were analyzed to determine the semantic orientation by using a lexicon-based approach. The lexicons used were the Bing Liu [22], Sentiment 140 [24], NRC [24], and Affin [1]. For each tweet, the strength of polarity (positivity and negativity) was computed and normalized according to the Euclidean norm or 2-norm, assuming that values near to zero are neutral:

$$\text{positivity} = \frac{\sum w_{+}}{\sqrt{(\sum w_{+})^2 + (\sum w_{-})^2}}$$

$$\text{negativity} = \frac{\sum w_{-}}{\sqrt{(\sum w_{+})^2 + (\sum w_{-})^2}}$$

Where

$w_{+}$ ; positive word weight in tweet.

$w_{-}$ ; negative word weight in tweet.

We defined positivity = 0 and negativity = 0 for neutral tweets, to avoid divisions by zero. In addition to the financial index polarity (*fip*) computed, according to each lexicon, we generated an information matrix conformed by the adjusting closing price of each financial index (*Price*), and conformed by vectors that contain the post date (*Date*), and each of the daily average lexicon values, calculated for each financial index (*S140*, *Bing<sup>H</sup>*, *Bing<sup>L</sup>*, *NRC<sup>H</sup>*, *NRC<sup>L</sup>*, *Afinn*).

$$fip_{NO,PQ}(\text{index}, \text{tweet}) = \begin{bmatrix} Date_T & Price_T & S140_T & Bing^H_T & Bing^L_T & NRC^H_T & NRC^L_T & Afinn_T \\ Date_I & Price_I & S140_I & Bing^H_I & Bing^L_I & NRC^H_I & NRC^L_I & Afinn_I \\ \vdots & \vdots \\ Date_v & Price_v & S140_v & Bing^H_v & Bing^L_v & NRC^H_v & NRC^L_v & Afinn_v \end{bmatrix}$$

In some cases, more than one tweet was published by a person or institution on the same date; so, we decided to average the strengths of the polarities of these tweets. Before computing these polarities, texts of tweets were pre-processed, removing numbers, punctuation symbols, and stop words.

We started from the assumption that the latent relationship between the financial indices' behavior and the polarity of sentiments in tweets can occur not always on the same day; that is, this relationship could be present a few days before a publication on Twitter, or a few days later.

To investigate the latent relationship, we computed the correlation of the adjusting closing prices, and the components of the financial index polarity (*fip*), defined above. The method is as follows: we first computed an offset or date shift (*s*) for the polarity vector. The date shifts can be forward or backward. In the first case (forward date shift), the offset is a positive integer (*shift* + +); in the second case (backward date shift), the offset is a negative one (*shift* - -). A value of zero in date shift means that the date of creation of a tweet is not modified. The correlation matrix ( $r(fip_{NO,PQ} index, tweet)$ ), based on the financial index polarity (*fip*), is calculated considering the coincidence of the dates between the adjusting closing prices and the components of the polarity vector with shifted dates (see Figure 1).

Finally, in order to identify the Twitter posts that had the greatest influence on the financial indices, we performed a search to find the value of the offset (from shift date = -7 to shift date = +7) that produced the most significant correlation in an absolute value:

$$\max_{NO,PQ} |r(fip_{NO,PQ} index, tweet)|$$

When there was not a coincidence of the dates between the adjusting closing prices and the components of the polarity vector with shifted dates; the correlation was defined as not available (NA).

## Results

The analysis of data was performed on a computer with the following characteristics. 2.2 GHz Intel Core i7 processor, 8 MB RAM, Mac OS 10.15.4. The method was implemented in the R programming language.

Table 4 shows an example of the polarity vectors (*Date*, *S140*, *BingH*, *BingI*, *NRCH*, *NRCI*, *Afinn*) found with data-set @business, related to Bloomberg Twitter account. A neutral polarity was determined with the S140 lexicon. Almost the same positive and negative polarity strengths were found with Bing's lexicon; while with the NRC one, a higher positive polarity strength than negative was determined. With the lexicon Affin a slightly negative polarity was computed. The shift date was equal to -7. This means that the original date of publications of the tweets were during May 5, 2020 and May 6, 2020 respectively, whereas the values of the adjusting close prices occurred during May 12, 2020, and May 13, 2020, respectively. The adjusting close prices corresponded to Shanghai, China's financial index.

Table 4. Example of polarity vector with a date shift = -7

Date	S140	Bing+	Bing-	NRC+	NRC-	Affin	Adj.Close
2020/05/12	0.000	0.293	0.257	0.4871	0.304	-0.028	2891.56
2020/05/13							
	0.000	0.231	0.338	0.510	0.382	-0.0379	2898.05

An example of the calculated correlations between the polarity and the adjusting closing price  $r(fip_{NO,PQ} index, tweet)$  is shown in Table 5; this example shows a perfect correlation between polarities and financial indices found in the @business data set.

Table 5. Correlations found for data set @business, date shift =-7

Date	S140	BingPos	BingNeg	NrcPos	NrcNeg	Affin
2020/05/12	NA	1	-1	-1	-1	1
2020/05/13	NA	-1	1	1	1	-1

After calculating the correlations of each financial indicator with the polarity generated by each lexicon ( $r(fip_{NO,PQ} index, tweet)$ ); we identify the best correlations to understand financial market behavior and sentiment analysis (Table 6). Not obtaining correlations for date shifts between -7 and 7 for dataset @NYTimes, we decided to extend this interval.

Table 6. Most significant correlations in absolute value for each data set

Data set	H1N1 (2009)			COVID-19 (2020)		
	Shift	max  r	lexicon	Shift	max  r	lexicon
		<i>shift date</i>			<i>shift date</i>	
@CNBusiness	-2	0.469	Affin	-6	0.704	Bing+
@Business	+2	0.540	S140	-7	1.000	Affin
@NYTimes	-1	0.398	S140	-13	0.973	S140
@Investingcom	NA	NA	NA	4	0.938	Affin

## Discussion

Sentiment analysis has the potential to be an important tool to improve decision-making in finance [5, 26], the analysis of polarity in Twitter has the potential to predict the financial market [29]. The information shared on Twitter affects investors' reactions, and therefore, affects stock indices, either positively or negatively, or in terms of upward or downward trends. Negative sentiments lead to lower risk tolerance [11].

One of the applications of understanding the emotions of social media consists of predicting the behavior of the financial indices diminishing the risk and uncertainty of the investors. Since investment portfolios can be formed with the shares that follow the tendency of the indices, diversifying with the selection of shares of different world markets and with the availability of more accurate and efficient sentiment measures and more wide-ranging studies [16].

Some of the analyzed Twitter accounts were unrelated to financial indices for several factors: 1) they had few followers, 2) they published very little, daily, or did not publish at all, 3) the publications were very specific and could affect a station or a sector, and therefore were not perceived in the index, 4) the Twitter account did not publish relevant information for the financial market; for these reasons, some Twitter accounts were not included in the correlation analysis.

High correlations were found in Investing, Bloomberg, and CNN Business accounts. One of the reasons they had high correlation was because they had a considerable number of followers: Investing 168,000, CNN Business 1.8 million, and Bloomberg 6.4 million and an average of 10 publications per day. Another Twitter account with high correlations was The New York Times (46.8 million followers) because it covered both financial and economic topics, such as music, culture, sports, art, and entertainment; which generated different kinds of sentiments in investors [33]; and it is a traditional means of communication among them.

We found that in 2009, the information posted on Twitter about H1N1 is practically non-existent; unlike COVID-19 in 2020, in this manner, our findings show that some of the Twitter posts had a more significant effect in the COVID-19 era, compared to H1N1. However, the effect took more days, compared to the moderate effect in 2009; this was because in 2020, the use of Twitter was intensive and more accounts published information about COVID-19 and its effects on finances.

The sentiment of investors could also predict the evolution of indices a few days in advance [33]. For the data of this investigation, the period in which the markets reacted, in the COVID-19 season, was 4 to 13 days after the information was shared and disseminated on Twitter; unlike the H1N1 season, which was 1-2 days [17].

Some studies reflect the importance of the investors' emotions in financial markets [8]. In this research, we confirmed the argument of Ahmed [2] in relation to the investors' reaction to a financial event, in the form of stock price movement, but applied to Twitter as there is an effect of the reactions generated by the Twitter posts on the financial indices days later. An important factor influencing the sentiments, found

on Twitter, is the sharp drop in the financial market caused by COVID–19, in addition to the fact that its duration and impact on the economy and health has been much greater, compared to H1N1.

We also confirm the argument of Sul and his colleagues [36], related to the number of followers on Twitter influences the performance of financial indices. Regardless of whether the publication is accurate or not, altering the feelings of a higher number of investors, forcing them to act in such a way that they seek the strategy that reduces risk. We verified that the Wall Street Journal [31] is an important source to know the behavior of the stock market. Moreover, we found three more sources: CNN News, Bloomberg, and Investing.com. We also found in our analysis that the lexicons that get the best results for sentiment analysis on Twitter were S140 and Affin.

## Conclusions

This investigation aims to study the effect of polarity on Twitter posts on world indices behavior in pandemic seasons. Most of the studies had been carried out under normal conditions; that is, they had not been analyzed in times of crisis, pandemics, or some disruptive economic event.

We proposed a correlation matrix that contains the polarity on Twitter and the financial market behavior, to study their relationship through the effect displaced, in days, by sentiment analysis and the adjusting close prices. Our findings show some important relationships days after the posts were made.

We found that the period in which the financial markets reacted was 6 to 13 days after the information was shared and disseminated on Twitter, in COVID–19 season; and 1 to 2 days, in H1N1 season. Still, we identified the inverse relation, where Twitter posted information about the financial market four days after.

Also, we found that in the 2009 pandemic, practically, there was no spread of H1N1 through Twitter and that several of the users considered influential did not have an account in 2009. More data and more information related to COVID was spread in 2020. The most influential Twitter accounts, in the COVID–19 season, were The New York Times, Bloomberg, CNN News, and Investing, presenting a very high relation between polarity on Twitter and stock market behavior.

We are aware that several factors influence market behavior. In this research, we found that the effect of social media publications was more significant in the times of COVID–19, compared to H1N1. First, we found that there were more active Twitter accounts in 2020. Second, there is virtually no published information about H1N1 in Twitter accounts that existed in 2009. Third, the drop in the rates, in the COVID–19 era, was more dramatic because there were more speculation, rumors, and negative news regarding the 2020 pandemic.

Fourth, there is an important effect of sentiments on Twitter on financial indices, a few days after their publication.

Finally, this research can be extended in two directions: First of all, the analysis of other indices, markets, or products, such as cryptosystems, FOREX, futures, etc. On the other hand, it will be necessary to analyze the reactions of investors with other emotional models that explain the behavior of the financial market. Besides, it is possible to analyze data from other social networks such as Facebook and YouTube.

## Declarations

*Conflict of Interest:* The authors declare that they have no conflict of interest.

## References

- [1] Affolter M, Slattery M, Mann RS. A lexicon for homeodomain-dna recognition. *Cell*, vol. 133, no. 7. pp. 1133–1135, 2008.
- [2] Ahmed SM, Quantification of investor emotion in financial news by analyzing the stock price reaction. In: 2017 International Conference on Information and Communication Technologies (ICICT). IEEE. 2017. pp. 119–123.
- [3] Atzeni M, Dridi A, Recupero DR. Using frame-based resources for sentiment analysis within the financial domain. *Progress in Artificial Intelligence*. Vol. 7, no. 4. pp. 273–294. 2018.
- [4] Cambria E. Affective computing and sentiment analysis. *IEEE Intelligent Systems*. Vol. 31. No. 2, pp. 102–107. 2016.
- [5] Cambria E, Hussain A. *Sentic computing: Techniques, tools, and applications*. Springer Science & Business Media. 2012, vol. 2.
- [6] Cambria E, Das D, Bandyopadhyay S, Feraco A. Affective computing and sentiment analysis. In: *A practical guide to sentiment analysis*. Springer, 2017, pp. 1–10.
- [7] Cambria E, Poria S, Gelbukh A, Thelwall M. Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*. Vol. 32, no. 6, pp. 74–80, 2017.
- [8] Chen H, De P, Hu HY, Hwang BH, “Sentiment revealed in social media and its effect on the stock market,” in 2011 IEEE Statistical Signal Processing Workshop (SSP). IEEE, 2011, pp. 25–28.
- [9] de la Cruz FM, Gomes MYFS. The influence of rumors in the stock market: A case study with Petrobras. *Transinforma,c~ao*, vol. 25, no. 3, pp. 187–193, 2013.
- [10] Day MY, Lee CC. Deep learning for financial sentiment analysis on finance news providers. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2016, pp. 1127–1134.
- [11] de la Orden MDC, Martínez RG, Vianez JP. Sentimiento de los medios de comunicación españoles en formato digital sobre el ibex 35. *aDResearch: Revista Internacional de Investigación en Comunicación*. No. 19, pp. 56–67, 2019.
- [12] Devitt A, Ahmad K. Sentiment polarity identification in financial news: A cohesion-based approach. In: *Proceedings of the 45th annual meeting of the association of computational linguistics*. 2007, pp. 984–991.
- [13] Atzeni M, Dridi A, Recupero DR. Fine-grained sentiment analysis on financial microblogs and news

- headlines. In: *Semantic Web Evaluation Challenge*. Springer. 2017, pp. 124–128.
- [14] Gómez-Martínez R, Prado-Román M, Plaza-Casado P. Big data algorithmic trading systems based on investors' mood. *Journal of Behavioral Finance*. Vol. 20, no. 2, pp. 227–238, 2019.
- [15] Hollander JB, Graves E, Renski H, Foster-Karim C, Wiley A, Das D. A (short) history of social media sentiment analysis. In: *Urban Social Listening*. Springer. 2016, pp. 15–25.
- [16] Kearney C, Liu S. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*. Vol. 33, pp. 171–185, 2014.
- [17] Kiyamaz H. The effects of stock market rumors on stock prices: evidence from an emerging market. *Journal of Multinational Financial Management*. Vol. 11, no. 1, pp. 105–115, 2001.
- [18] Kraus M, Feuerriegel S. Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. *Expert Systems with Applications*. Vol. 118, pp. 65–79, 2019.
- [19] Liew J, Budavari T. The “sixth” factor—a social media factor derived directly from tweet sentiments. *The Journal of Portfolio Management*. Vol. 43, no. 3, pp. 102–111, 2017.
- [20] Liew JKS, Budavári T. Do tweet sentiments still predict the stock market. Available at SSRN 2820269. 2016.
- [21] Liu B. Many facets of sentiment analysis. In: *A practical guide to sentiment analysis*. Springer. 2017, pp. 11–39.
- [22] Liu B. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.
- [23] López-Chau A, Valle-Cruz D, Sandoval-Almazán R. Sentiment analysis of twitter data through machine learning techniques. In: *Software Engineering in the Era of Cloud Computing*. Springer. 2020, pp. 185–209.
- [24] Mohammad SM, Turney PD. *NRC emotion lexicon*. National Research Council. Canada, 2013.
- [25] Poria S, Cambria E, Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*. Vol. 108, pp. 42–49, 2016.
- [26] Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*. Vol. 37, pp. 98–125, 2017.
- [27] Poria S, Cambria E, Howard N, Huang GB, Hussain A. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*. Vol. 174, pp. 50–59, 2016.
- [28] Ruiz-Martínez JM, Valencia-García R, García-Sánchez F. Semantic-based sentiment analysis in financial news. In: *Proceedings of the 1st International Workshop on Finance and Economics on the Semantic Web*, 2012, pp. 38–51.
- [29] Smailović J, Grčar M, Lavrač N, Žnidaršič M. Predictive sentiment analysis of tweets: A stock market application. In: *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*. Springer, 2013, pp. 77–88.
- [30] Sohagir S, Wang D, Pomeranets A, Khoshgoftaar TM. Big data: Deep learning for financial sentiment analysis. *Journal of Big Data*. Vol. 5, no. 1, p. 3, 2018.
- [31] Souza TTP, Kolchyna O, Treleven PC, Aste T. Twitter sentiment analysis applied to finance: A case study in the retail industry. arXiv preprint arXiv:1507.00784, 2015.
- [32] Sun A, Lachanski M, Fabozzi FJ. Trade the tweet: Social media text mining and sparse matrix

factorization for stock market prediction. International Review of Financial Analysis. Vol. 48, pp. 272–281, 2016.

[33] Zhang W, Li X, Shen D, Teglio A. Daily happiness and stock returns: Some international evidence. Physica A: Statistical Mechanics and its Applications. Vol. 460, pp. 201–209, 2016.

[34] Zhao L, Li L, Zheng X. A bert based sentiment analysis and key entity detection approach for online financial texts. arXiv preprint arXiv:2001.05326, 2020.

[35] Yang SY, Mo SYK, Liu A. Twitter financial community sentiment and its predictive relationship to stock market movement. Quantitative Finance. Vol. 15, no. 10, pp. 1637–1656, 2015.

[36] Sul HK, Dennis AR, Yuan L. Trading on twitter: Using social media sentiment to predict stock returns. Decision Sciences. Vol. 48, no. 3, pp. 454–488, 2017

## Figures

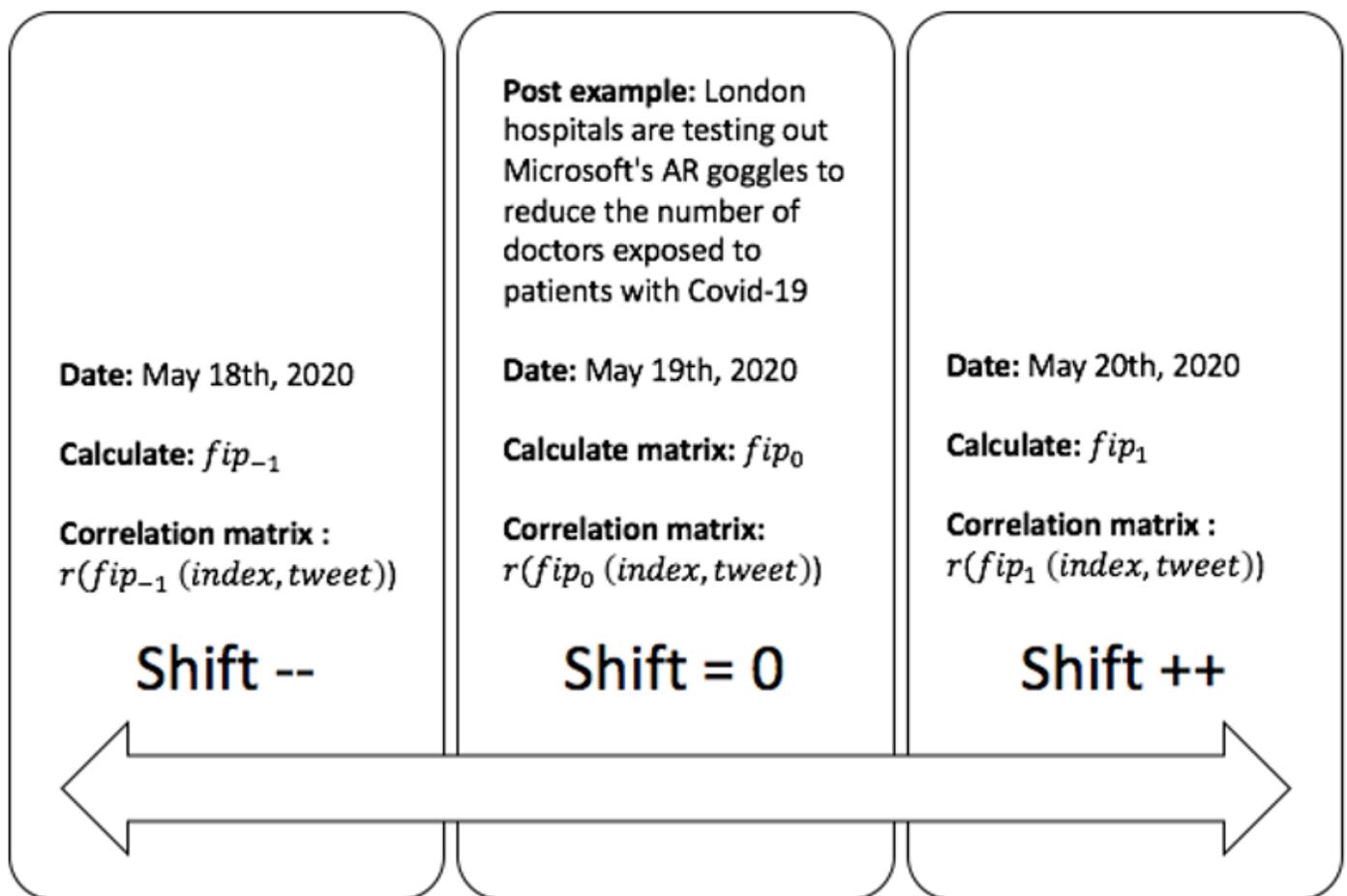


Figure 1

Correlation matrix with shifting calculation

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [APPENDICES.docx](#)