



# Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods

David Valle-Cruz<sup>1</sup> · Vanessa Fernandez-Cortez<sup>2</sup> · Asdrúbal López-Chau<sup>3</sup> · Rodrigo Sandoval-Almazán<sup>4</sup>

Received: 15 June 2020 / Accepted: 5 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

Investors are constantly aware of the behaviour of stock markets. This affects their emotions and motivates them to buy or sell shares. Financial sentiment analysis allows us to understand the effect of social media reactions and emotions on the stock market and vice versa. In this research, we analyse Twitter data and important worldwide financial indices to answer the following question: How does the polarity generated by Twitter posts influence the behaviour of financial indices during pandemics? This study is based on the financial sentiment analysis of influential Twitter accounts and its relationship with the behaviour of important financial indices. To carry out this analysis, we used fundamental and technical financial analysis combined with a lexicon-based approach on financial Twitter accounts. We calculated the correlations between the polarities of financial market indicators and posts on Twitter by applying a date shift on tweets. In addition, correlations were identified days before and after the existing posts on financial Twitter accounts. Our findings show that the markets reacted 0 to 10 days after the information was shared and disseminated on Twitter during the COVID-19 pandemic and 0 to 15 days after the information was shared and disseminated on Twitter during the H1N1 pandemic. We identified an inverse relationship: Twitter accounts presented reactions to financial market behaviour within a period of 0 to 11 days during the H1N1 pandemic and 0 to 6 days during the COVID-19 pandemic. We also found that our method is better at detecting highly shifted correlations by using SenticNet compared with other lexicons. With SenticNet, it is possible to detect correlations even on the same day as the Twitter posts. The most influential Twitter accounts during the period of the pandemic were The New York Times, Bloomberg, CNN News and Investing.com, presenting a very high correlation between sentiments on Twitter and stock market behaviour. The combination of a lexicon-based approach is enhanced by a shifted correlation analysis, as latent or hidden correlations can be found in data.

**Keywords** Sentic computing · Sentiment analysis · Affective computing · Finance · Stock market · Pandemic

---

✉ David Valle-Cruz  
davacr@uaemex.mx; davacr@gmail.com

Vanessa Fernandez-Cortez  
vfernandezc@uaemex.mx

Asdrúbal López-Chau  
alchau@uaemex.mx

Rodrigo Sandoval-Almazán  
rsandovala@uaemex.mx

<sup>2</sup> Department of Finance, Autonomous University of the State of Mexico, Instituto Literario 100, Toluca, Mexico

<sup>3</sup> CU UAEM Zumpango, Autonomous University of the State of Mexico, KM 3.5, Camino Viejo a Jilotzingo, Valle Hermoso, 55600 Zumpango, Estado de México, Mexico

<sup>4</sup> Department of Political Science, Autonomous University of the State of Mexico, Instituto Literario 100, Toluca, México

<sup>1</sup> Department of Engineering, Autonomous University of the State of Mexico, Instituto Literario 100, Toluca, Mexico

## Introduction

In 2020, the world was rocked by the COVID-19 pandemic, and global concerns related to two major issues arose: health and the economy. Stock markets around the world collapsed under the uncertainty generated by the SARS-CoV-2 virus, and something similar happened with the H1N1 pandemic. In this regard, different studies have been conducted in attempts to understand the effects of pandemics in different areas. In particular, this article aims to study the relationship between the sentiments generated on Twitter and the stock market through sentic computing techniques.

Sentic computing is a recent line of research that is supported by computational techniques, neuroscience, sociology, psychology and mathematics that have the potential to be applied to different areas of knowledge [1]. Some sophisticated artificial intelligence techniques allow for the analysis of different types of data, whether text, images, audio, or video, to determine the sentiment, emotions, or meaning of the content of big data [2]. Areas such as computer science, mathematics, neuroscience, business, administration, accounting, engineering, arts and humanities, earth sciences, psychology, economics, econometrics and finance make use of and are complemented by sentiment analysis [3].

Additionally, stock market prediction is one of the most challenging problems and has concerned both researchers and financial analysts for more than half a century [4, 5]. Research on this topic during pandemics is practically non-existent. However, Khatua and colleagues [6] studied the 2014 Ebola and 2016 Zika outbreaks, finding that smaller domain-specific input corpora from the Twitter corpus are better in extracting meaningful semantic relationships than generic pre-trained Word2Vec or GloVe. In this regard, financial markets are based on expectations [7], and investors buy or sell shares depending on the sentiments of fear or desire caused by some event [8]. Investors bet on the credibility of the media outlets that publish rumours despite knowing that the information is not always reliable [9, 10]. However, this trend is growing with the use of online social media networks such as Twitter, Facebook, Instagram and YouTube [11]. Exposure to topics and ideas is thus increased through these platforms, affecting the stock market and investors' emotions. Emergencies such as earthquakes, pandemics, political events and terrorism are special cases where the amount of available information is reduced, and decisions have to be made in uncertain circumstances [12].

Indeed, the COVID-19 pandemic has changed our lives, altered the balance within stock markets around the world and created an unstable environment for at least several weeks depending on the country [13]. A similar situation

arose 11 years ago, with the pandemic caused by the H1N1 virus [14]; these kinds of events affect the behaviour, emotions, sentiments and feelings of investors. In this regard, emotions and sentiments are inherent to investors' reactions and decision-making, learning, communication and awareness [15]. Understanding emotions is one of the most important aspects of personal development and growth and is paramount to the emulation of human intelligence. The processing of emotions is important for the advancement of artificial intelligence and a task closely related to the detection of polarity and emotions [16]. The potential for this type of artificial intelligence technique has extended to different fields, such as social media and finance, where sentic computing has been applied to find patterns of behaviour, meaning identification, emotions, trends and even the induction of people's decisions.

Research on sentic computing and finance has grown, but very few studies have analysed social media datasets during pandemics. It is important to identify signs or patterns that guide investors when selecting the assets that will constitute their portfolios. Investors try to ensure that the prices of selected stocks have a trend that is either very similar to or higher than the corresponding indices. Therefore, an investor must have tools that allow him/her to factor in the most considerable information available for technical and fundamental analysis. This leads him/her to reduce risk and uncertainty as much as possible, as well as maximize returns [17].

The information shared on social media generates sentiments and reactions in investors that influence their decisions to buy or sell financial assets, especially on the stock market [18]. News shared on social media, whether true or not, cause changes in the trends of international stock market indices [19]. Some studies confirm the relationship between sentiments generated by social media in investors and market trends, and the current study continues this path in exploring pandemic periods with a lexicon-based approach to detect polarity in financial news on Twitter. This study aims to address this gap in literature. To fulfil this aim, the study seeks to compare similar emergencies, as well as the use of social media platforms that improve the dissemination of rumours and news that alter investors' perceptions of the stock exchange. Although there are different techniques and applications for computational finance, no one knows for how long an event will affect the financial market, specifically whether posts on social media, such as Twitter, affect financial indicators and vice versa. This is especially true during pandemics such as H1N1 and COVID-19. To achieve this, we analysed shifting correlations of polarity generated by Twitter accounts that influence the financial sector and its relationship with the most important financial indices worldwide.

The following question guides the study: *How does the polarity generated by Twitter posts influence the behaviour of financial indices during pandemics?* To answer this question, our analysis proposes a methodology based on the financial sentiment analysis of influential Twitter accounts in the financial field and shifted correlations with the behaviour of important financial indices. The document is divided into five sections. In the first section, we present the state of the art related to sentic computing and finance. In the second section, we show our methodology, which is based on fundamental and technical financial analysis combined with a lexicon-based approach on financial Twitter accounts and shifted correlations during pandemics. The third section describes the findings of the shifted correlations between the polarities of financial market indicators and posts on Twitter. The fourth section is the discussion of results. Finally, we present the conclusions of the study and suggestions for future work.

## State of the Art

This section consists of two stages. First, we present a brief introduction to sentic computing; the second section describes the state of the art of sentic computing and finance based on the 13 scientific papers existing in Google Scholar, Scopus, Web of Science, IEEE Xplore, Springer Link and ACM Digital Library.

### A Brief Introduction to Sentic Computing

Emotions are fundamental for successful and effective communication between human beings [20, 21]. In fact, in many everyday situations, emotional intelligence is more important than the intelligence quotient (IQ) for successful interactions [16], and social media is a massive modern source for detecting people's emotions and sentiments [22]. For instance, content posted on Facebook and Twitter shows the opinions, emotions and sentiments of people about any kind of event [23].

Affective computing and sentiment analysis are multi-disciplinary areas that include academics and professionals from different disciplines, such as computer science, cognitive science, political science, economics and finance [24–27]. In particular, sentiment analysis is an area of artificial intelligence that allows us to extract and analyse data stored on social media in images, sounds and videos. Sentiment analysis allows the identification of patterns or characteristics in long data sets. This can be very useful for decision-making in organizations, political movements, business strategies, marketing campaigns and product preferences, among others [1, 15, 24]. Sentiment analysis reveals personal opinions towards entities such as

products, services, or events. This helps organizations and companies to improve their marketing, communication, production and acquisition [28].

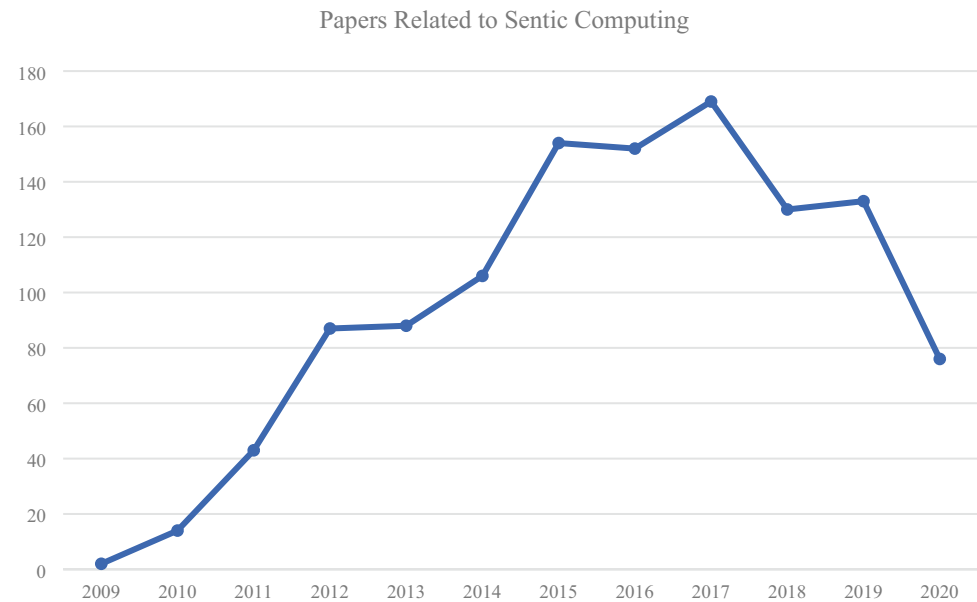
Cambria and colleagues [16] argued that sentiment analysis has been confused with the task of polarity detection (detecting negative and positive sentiments). However, this is only one of many natural language processing (NLP) problems that must be solved to achieve human-like performance in sentiment analysis. Thus, when performing sentiment analysis, we encounter a complex problem of NLP that can be analysed by different emotional or psychological models [1, 29].

The opportunity to automatically capture the sentiments of the public regarding social events, political movements, marketing campaigns and product preferences has led to growing interest both in the scientific community and in the business world. The scientific community is attracted to exciting open challenges, and the business world is attracted based on the notable consequences for marketing and the incorporation of sentiments into financial market predictions [29, 30]. This has led to the emergence of the fields of affective computing and sentiment analysis, which take advantage of human–computer interactions, information retrieval and multimodal signal processing to identify people's emotions from the growing amount of social data online [31].

In this regard, the main goal of sentic computing is to create emotion detection models supported by different disciplines. It proposes the application of artificial intelligence and semantic web techniques for the representation of knowledge, mathematics for making inferences, linguistics for analysing discourse and pragmatics, psychology for cognitive and affective modelling, sociology for understanding the dynamics of social networks and social influence and ethics for understanding the nature of the mind [1].

The first wave of sentic computing identified as the state of the art dates from 2009. Cambria, Hussain, Havasi and Eckl [32] published a seminal study related to the detection of emotions in texts for the development of intelligent systems called AffectiveSpace. Today, there are 1154 documents on the Internet regarding sentic computing and the second version of AffectiveSpace, a vector space model that allows for reasoning by analogy on natural language concepts [25]. In 2017, for instance, 169 papers were published on this topic (Fig. 1).

On Google Scholar, there are 876 papers, including reports and conference proceedings. Regarding scientific databases such as Scopus, Web of Science, IEEE Xplore, Springer Link and ACM Digital Library, there are a total of 278 documents, including articles, book chapters, books and proceedings from international conferences related to the application of sentic computing in different areas (Table 1).

**Fig. 1** Sentic computing papers per year

## Sentic Computing and Finance

It is widely accepted that public mood is correlated with financial markets. In this regard, sentic computing has great potential to improve the capacity of customer relationship management and referral systems [25]. For example, sentic computing helps to reveal what characteristics clients enjoy or which items to exclude from referrals based on what previously received a negative response. Business and financial intelligence are also major factors in companies' interest in affective computing and sentiment analysis [16].

Investors have always been interested in stock price forecasting. However, little research has examined the effects of social media on forecast results specifically during

pandemics. In this regard, Day and Lee [33] argued that financial sentiment analysis is an important area of financial technology research (FinTech).

Ahmed [34] stated that investor reactions to a financial event, in the form of stock price movement, reflect the severity of the event. He also argued that the words used in published news show a similar degree of emotion.

Computational techniques applied to financial analysis are related to natural language processing, sentiment analysis, text mining, artificial neural networks, machine learning, biologically inspired computational approaches and vine-growing algorithms. Regarding the literature on sentic computing and finance, we identified 13 scientific papers that show the application of financial sentiment analysis (Table 2).

**Table 1** Sentic computing papers on the Internet

Year	Google Scholar	Scopus	Web of Science	IEEE Xplore	Springer Link	ACM Digital Library
2009	2	0	0	0	0	0
2010	8	0	3	1	2	0
2011	26	4	4	0	8	1
2012	45	11	6	2	22	1
2013	60	6	4	4	11	3
2014	83	7	6	1	8	1
2015	116	6	5	2	23	2
2016	126	6	6	1	12	1
2017	137	4	8	1	17	2
2018	106	2	2	1	19	0
2019	105	4	3	1	19	1
2020	62	1	0	2	11	0

**Table 2** Scientific papers related to sentic computing and finance

Reference	Title
Minhas et al. (2013) [35]	A Review of Artificial Intelligence and Biologically Inspired Computational Approaches to Solving Issues in Narrative Financial Disclosure
Ahmed (2017) [34]	Quantification of Investor Emotion in Financial News by Analyzing the Stock Price Reaction
Atzeni, Dridi and Recupero (2018) [36]	Using Frame-Based Resources for Sentiment Analysis within the Financial Domain
Xing, Cambria and Welsch (2018) [37]	Intelligent Asset Allocation via Market Sentiment Views
Picasso et al. (2018) [38]	Technical Analysis and Sentiment Embeddings for Market Trend Prediction
Xing, Cambria and Welsch (2018) [39]	Natural Language Based Financial Forecasting: a Survey
Malandri et al. (2018) [40]	Public Mood-Driven Asset Allocation: the Importance of Financial Sentiment in Portfolio Management
Upreti et al. (2019) [41]	Knowledge-Driven Approaches for Financial News Analytics
Dridi, Atzeni and Recupero (2019) [42]	FineNews: Fine-Grained Semantic Sentiment Analysis on Financial Microblogs and News
Merello et al. (2019) [43]	Ensemble Application of Transfer Learning and Sample Weighting for Stock Market Prediction
Xing, Cambria and Welsch (2019) [44]	Growing Semantic Vines for Robust Asset Allocation
Xing, Cambria and Zhang (2019) [45]	Sentiment-Aware Volatility Forecasting
Akhtar, Ekbal and Cambria (2020) [46]	How Intense Are You? Predicting Intensities of Emotions and Sentiments Using Stacked Ensemble

The first study on the state of the art of sentic computing and finance was conducted in 2013 using a biologically inspired computational approach to solve the problems of narrative financial disclosures [35]. In 2017, Ahmed [34] proposed a lexicon useful to identify investor emotions in financial news. He assumed that the words used in the news to describe an event also exhibit a similar degree of emotion. In 2018, Xing, Cambria and Welsch [39] published a study of articles that harness NLP techniques to make financial market predictions by identifying the research field of natural language-based financial forecasting (NLFF).

Atzeni, Dridi and Reforgiato [36] considered that user-generated data in blogs and on social media have become a valuable resource for sentiment analysis in the financial domain. Therefore, they suggested a fine-grained approach that returns a continuous score in the range of  $[-1, +1]$  to identify polarity. Malandri and colleagues [40] investigated whether public mood ascertained from social media and online news is correlated with or predictive of portfolio returns, introducing the framework of sentiment-driven portfolio allocation. This novel approach automatically produces an optimal online portfolio allocation strategy.

Xing, Cambria and Welsch [37] investigated the role of market sentiment in asset allocation problems. They proposed the computation of sentiment time series from social media data, finding that the sentiment time series obtained through sentic computing is comparable to that obtained through some commercial tools. Picasso and colleagues [38] combined both technical and fundamental analysis through the application of data science and machine learning

techniques. They found a robust predictive model able to forecast the trend of a portfolio composed of the twenty most capitalized companies listed in the NASDAQ100 index.

Also in 2019, Xing, Cambria and Welsch [44] proposed leveraging prior semantic knowledge of assets to find a suitable vine structure for financial portfolio optimization. Their findings suggest that the construction of a semantic vine improves on the arbitrary vine-growing method in the context of robust correlation estimation and multi-period asset allocation. Merello and colleagues [43] showed that the return predictions generated by a regression approach are more meaningful concerning the “buy” or “sell” signals provided by classification approaches during trading. They found that the application of transfer learning and sample weighting over different market fluctuations has been instrumental in enhancing performance. This is especially true for the largest and most important returns. Xing, Cambria and Zhang [45] proposed a novel model termed sentiment-aware volatility forecasting (SAVING). This model incorporates market sentiment for stock return fluctuation prediction and aims to provide a more accurate estimation of the temporal variances of asset returns by better capturing the bidirectional interaction between movements of asset prices and market sentiment. Upreti and colleagues [41] provided an overview of recent developments in the domain of financial news analytics, with a special focus on sentiment analysis and event detection. They observed that robo-readers and sentiment analysis techniques capable of making automated decisions are still in their infancy.

Dridi, Atzeni and Recupero [42] proposed a supervised method and found that semantic features and semantic frames



can be applied successfully to sentiment analysis within the financial domain, thus leading to better results. Finally, Akhtar, Ekbal and Cambria [46] proposed a stacked ensemble method for predicting the degree of intensity for emotion and sentiment by combining the outputs obtained from several deep learning and classical feature-based models using a multi-layer perceptron network. To achieve this, Akhtar and colleagues focused on emotion analysis in the generic domain and sentiment analysis in the financial domain. Their model achieved better performance than existing state-of-the-art systems.

In summary, the application of sentic computing in finance has focused on sentiment analysis generated by news related to the stock market. Other applications have focused on the prediction of the behaviour of financial investment portfolios and the detection of useful patterns for decision-making (Fig. 2).

## Methodology

This section describes the methodology of the study based on the comparison of different lexicons to determine the polarity of Twitter posts and their relation to the behaviour of financial indices. This section is divided into two stages. First, we present the fundamental and technical financial analysis of the stock market and Twitter data. Second, we describe the process of extracting and analysing data based on the lexicon approach and date-shifted correlations.

## Fundamental and Technical Analyses

The methodology is based on sentiment analysis applied to Twitter accounts that influence stock market behaviour. To achieve the study objective, we downloaded data from Twitter at two important moments: during the H1N1 pandemic and during the COVID-19 pandemic. For the H1N1 pandemic, the period for data collection was from June to July 2009. For the COVID-19 pandemic, the period for data collection was from January to May 2020. The above periods were selected based on the time when the first stock market index peaked and began to fall, i.e., the point at which the pandemics started to influence financial indices. Pandemics cause prices to fall to a minimum and lead to a period of financial disruption and uncertainty. Since the COVID-19 pandemic has been the most damaging to the financial market, we extended the Twitter data analysis period to May 2020.

We selected the financial indices because they are the most representative and influential for each continent. Additionally, we included companies from different sectors, which provided a general perspective of their performance and behaviour. The financial data used for this study consisted of adjusted closing prices, where each of the financial indices in the world reached their maximum and minimum. This was first done for the H1N1 influenza period and then for the COVID-19 period. The analysed stock market indices were IPC (in Spanish, Índice de Precios y Cotizaciones), S&P 500 (Standard

**Fig. 2** Word cloud of scientific papers related to sentic computing and finance



**Table 3** Description of financial indicators

Country	Index	Stock market	Involved sectors
Mexico	IPC	Mexican Stock Exchange	Telecommunications, consumer products, materials, financial services, industrial, healthcare, financial services
USA	NASDAQ 100	NASDAQ	Telecommunications, biotechnology, video games, hardware, software, Internet services, financial services, pharmaceuticals, E-commerce, data analysis
USA	Dow Jones	New York Stock Exchange and NASDAQ	Covers all sectors except transport and utilities
USA	S&P 500	New York Stock Exchange and NASDAQ	Consumer discretionary, health care, industrials, information technology, materials, real state, communication services, utilities, financials, energy
UK	FTSE 100	London Stock Exchange	Financial services, mining, food, pharmaceutical and technology, software and computer services, life insurance, services
Brazil	BOVESPA	Sao Paulo Stock Exchange	Energy, logistics, finance and insurance, real estate, energy, steel and metallurgy, aviation, telecommunications, food
France	CAC 40	Euronext	Energy, logistics, banking, real estate, energy, oil and gas, aerospace, telecommunications, food, oil equipment, industrial machinery
Germany	DAX	Euronext	Logistics, finance and insurance, real estate, energy, pharmaceutical chemicals, airlines, telecommunications, consumer goods, automotive, utilities, building materials, technology, software
Japan	Nikkei 225	Tokyo 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 and insurance, petrochemical, food

& Poor's 500), NASDAQ 100 (National Association of Securities Dealers Automated Quotation 100), Dow Jones, FTSE 100 (Financial Times Stock Exchange 100), BOVESPA (in Portuguese, Bolsa de Valores do Estado de São Paulo), CAC 40 (in French, Cotation Assistée en Continu), DAX (in German, Deutscher Aktienindex), Hang Seng, Nikkei 225 (Nikkei Heikin Kabuka 225) and SSE Composite (Shanghai Stock Exchange Composite) (Table 3).

Once the dates for the maximum and minimum prices were identified in the stock market behaviour, the prices from January 2009 to May 2020 were downloaded from the Yahoo Finance site and Investing.com, which is a common practice of stock market analysts. Then, the percentage loss of each of the indices was calculated (Table 4). Adjusted closing prices were selected because they represented the closing price and the corporate earnings to which each one was entitled once the share was acquired. The adjusted closing price was used when analysing historical returns and was calculated in terms of the stock exchange to which it belonged.

We selected Twitter posts from financial markets, finance and the economy for this study, collecting data from media accounts recommended by financial experts

[47]. Another source of data was Internet sites that published general and international news that affected the world's population and therefore affected the sentiment of investors. Twitter accounts selected for sentiment analysis consisted of personal, company, or organizational accounts, as well as news broadcasts that were influential or important for finance. Some financial influencers do not have a Twitter account; for this reason, we downloaded Tweets from the company or organization they run, when available. Another important issue is that some Twitter users had an account in 2020 but not in 2009 due to the variation in the adoption and use of social media (NA\*) (Table 5).

### Data Extraction and Analysis

Tweets were analysed to determine semantic orientation by using a lexicon-based approach. In this approach, texts were tokenized by splitting the text into individual words. The stemming of the words was then computed. Next, stop words, punctuation symbols, numbers and other special characters were removed. A lexicon containing both words and a measure of polarity was used to infer emotion in texts. Lexicon-based analysis is a very simple technique

**Table 4** Maximum and minimum prices of the stock exchange indices

Index	Maximum price date	Maximum price	Minimum price date	Minimum price	Percentage loss (%)
During the H1N1 pandemic					
IPC	11/June/2009	25,372.84	07/July/2009	23,359.93	− 7.93
NASDAQ 100	11/June/2009	1862.36	07/July/2009	1746.17	− 6.23
Dow Jones	11/June/2009	8770.91	10/July/2009	8146.52	− 7.11
S&P 500	11/June/2009	944.89	10/July/2009	879.13	− 6.95
FTSE 100	10/June/2009	4436.80	10/July/2009	4127.20	− 6.97
BOVESPA	12/June/2009	53,558.00	14/July/2009	48,873.00	− 8.74
CAC 40	11/June/2009	3334.93	10/July/2009	2983.10	− 10.55
DAX	11/June/2009	5107.25	10/July/2009	4576.31	− 10.39
Hang Seng	11/June/2009	18,791.02	13/July/2009	17,254.63	− 8.17
Nikkei 225	12/June/2009	10,135.82	13/July/2009	9050.33	− 10.70
SSE Composite	10/June/2009	2816.25	12/June/2009	2743.76	− 2.57
During the COVID-19 pandemic					
IPC	12/February/2020	45,338.37	23/March/2020	32,964.21	− 27.29
NASDAQ 100	19/February/2020	9,817.17	20/March/2020	6,879.52	− 29.92
Dow Jones	12/February/2020	29,551.41	23/March/2020	18,591.92	− 37.08
S&P 500	19/February/2020	3386.14	23/March/2020	2237.39	− 33.92
FTSE 100	19/February/2020	7457.02	20/March/2020	5190.78	− 30.39
BOVESPA	19/March/2020	68,332.00	20/March/2020	67,069.00	− 1.84
CAC 40	19/February/2020	6111.24	18/March/2020	3754.84	− 38.55
DAX	19/February/2020	13,789.00	18/March/2020	8441.70	− 38.77
Hang Seng	17/February/2020	27,959.59	23/March/2020	21,696.13	− 22.40
Nikkei 225	12/February/2020	23,861.21	19/March/2020	16,552.83	− 30.62
SSE Composite	13/January/2020	3115.57	23/March/2020	2660.17	− 14.61

to implement, and it yields good results in most cases. In this paper, data (posts on Twitter) were pre-processed according to the steps shown in Algorithm 1.

We then analysed the cleaned data by applying the lexicon-based approach.

**Table 5** Analysed 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 in 2009

### Lexicon-Based Analysis

We used the following lexical items to determine the sentiment of a text: Bing Liu [26], Sentiment 140 [48], NRC [49], Affin [50] and SenticNet [51]. Each of these lexicons has—as part of their structure—the elements word and value. The former (word) is a list of terms that are commonly used on social networks. The latter (value) is a measure of polarity that corresponds to the word used. Bing Liu's Opinion Lexicon uses the values “positive” and “negative” to rate words. Sentiment 140 classifies the polarity of words as follows: 0 for negative, 2 for neutral and 4 for positive. The NRC lexicon uses a set of sentiments instead of numeric values to determine the polarity of words. The sentiments used are trust (+), fear (-), negative (-), sadness (-), anger (-), surprise(+), positive(+), disgust (-), joy (+) and anticipation (+). In this set, the symbols “-” and “+” refer to negative and positive polarity, respectively. In the Affin lexicon, terms are rated for valence with an integer between − 5 (most negative polarity) and + 5 (most



**Algorithm 1** Preprocessing of Data**Input:** Tweets downloaded**Output:** Data cleaned.**For** each tweet **do****Step 1:** Split text**Step 2:** Remove punctuation symbols, interrogation and exclamation symbols**Step 3:** Remove special elements, such as hashtags and html symbols**Step 4:** Remove stop words**Step 5:** Compute the stem of remaining words**Step 6:** Join the words to form a string

positive polarity). SenticNet provides an API that allows for the easy application of sentic computing. We used this API, but the polarity of each word was computed by invoking the `polarity_intense()` method. During analysis, the polarity values were transformed into numerical values. This process is shown in Table 6.

**Table 6** Transformation of values of polarity into real numbers

Lexicon	Rule
Bing Liu	Value "positive" is transformed into + 1 Value "negative" is transformed into - 1
Sentiment 140	Value 0 is transformed into - 1 Value 2 is transformed into 0 Value 4 is transformed into + 1
NRC	Sentiments (+) are transformed into + 1 Sentiments (-) are transformed into - 1
Affin	Values are divided by 5
SenticNet	<code>polarity_intense ()</code> method

Algorithm 2 shows the procedure used to retrieve the value of the polarity intensity of a text.

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 zero are neutral (see Eqs. (1) and (2)):

$$\text{positivity} = \frac{\sum w_{pi}}{\sqrt{(\sum w_{pi})^2 + (\sum w_{ni})^2}} \quad (1)$$

$$\text{negativity} = \frac{\sum pw_{ni}}{\sqrt{(\sum w_{pi})^2 + (\sum w_{ni})^2}} \quad (2)$$

where

$w_{pi}$ : positive weight of  $i$ th word in text  $\mathbf{T}$ .

$w_{ni}$ : negative weight of  $i$ th word in text  $\mathbf{T}$ .

We defined  $\text{positivity} = 0$  and  $\text{negativity} = 0$  for neutral tweets, which allowed us to avoid division by 0.

In some cases, more than one tweet was published by a person or institution on the same date; for this reason, we decided to average the polarities of these tweets.

**Algorithm 2** Calculating Polarity Intensity**Input:** Text  $\mathbf{T}$ , Lexicon  $\mathbf{L}$ **Output:** Polarity intensity of  $\mathbf{T}$ **Step 1:** Split  $\mathbf{T}$  into a list  $\mathbf{W}$  of words**For** each  $w \in \mathbf{W}$  **do****Step 2:** Recover *polarity intensity of  $w$ : value of term  $t \mid t \in \mathbf{L}, t = w$* **Step 3:** Convert *polarity* to a real number (see Table 6)**Step 4:** If polarity is positive, then assign  $w_{pi} \leftarrow \text{polarity}$ **Step 5:** Else assign  $w_{ni} \leftarrow \text{polarity}$ **Step 6:** Compute positivity and negativity of  $\mathbf{T}$  (see equations (1) and (2), respectively)

**Table 7** Hypothetical example of a matrix  $M_{pricesentic}$  of dates, adjusted prices and polarities of posts

Date	AdjPrice	S140+	S140-	Bing+	Bing-	NRC+	NRC-	Affin+	Affin-	SN+	SN-
12-March-2020	2923.49	- 0.56	0.24	0.33	- 0.14	0.12	- 0.16	0.58	- 0.06	0.05	0.57
13-March-2020	2887.43	0.10	0.31	0.34	- 0.18	0.50	- 0.20	0.44	- 0.41	0.70	- 0.52
16-March-2020	2789.25	0.50	0.28	- 0.07	- 0.59	0.14	0.33	0.52	0.52	- 0.03	0.44

### Date-Shifted Correlations

We computed the correlation between the financial indices and polarities of posts published on Twitter. The value of the correlation was used as an indicator of the latent relationship between the behaviours of these two variables. The relationships between the financial indices on day **A** and the polarity of sentiments in tweets published on day **B** can be discovered by applying a date shift on day **B**. This allowed the detection of correlations presented a number of days before or after a Twitter post.

To identify whether the Twitter posts made on a given date may have had any correlation with the financial indices reported on a date different from the post, we decided to simulate a (mathematical) translation of the date of the post. In a translation, all dates of posts were moved (forward or

$S140^+$ ,  $S140^-$ : positivity and negativity of a tweet according to S140's lexicon, respectively.

$Bing^+$ ,  $Bing^-$ : positivity and negativity of a tweet according to Bing Liu's lexicon, respectively.

$NRC^+$ ,  $NRC^-$ : positivity and negativity of a tweet according to the NRC lexicon, respectively.

$Affin^+$ ,  $Affin^-$ : positivity and negativity of a tweet according to Affin's lexicon, respectively.

$SN^+$ ,  $SN^-$ : polarity computed with SenticNet.

Since more than one text was usually posted by the same Twitter user, these form a matrix  $M_{pricesentic}$  of vectors **V**. Table 7 shows a hypothetical example of such a matrix formed with three texts published on different dates.

The correlations corresponding to the columns of positivity with an adjusted price and those of negativity with an adjusted price are calculated. Algorithm 3 shows this procedure:

---

#### Algorithm 3 Detecting correlations with date shift

---

**Input:** *Index*, *date-shift* value

**Output:** Maximum value of correlation  $\left( \max_{shift-date} |r(M_{pricesentic})| \right)$

**Step 1:** Apply date shift to the dates of posts.

**Step 2:** Compute  $M_{pricesentic}$  corresponding to *index*

**Step 3:** Compute the correlation between column **AdjPrice** and polarities

**Step 4:** Search for the maximum absolute value of correlations (see equation (4))

---

backward) to the same number of days in the same direction. We called this translation date shift.

The result provided a shifted value. Based on the price of a financial index reported on  $Date_i$  and a text of a tweet created on  $Date_i + shift$ , we built a vector **V** with the components shown below:

To identify the Twitter posts that had the greatest correlation with the financial indices, we performed a search to find the value of the offset that produced the most significant correlation in an absolute value (see Eq. (3)):

$$\max_{shift-date} |r(M_{pricesentic})| \quad (3)$$

$$V = Date_i + shift, Adjprice_i, S140^+, S140^-, Bing^+, Bing^-, NRC^+, NRC^-, Affin^+, Affin^-, SN]$$

where:

**shift**: number of days to translate  $Date_i$ , forward if  $shift > 0$ , backward if  $shift < 0$ , no date translation otherwise.

**Price**: the adjusted closing price of the financial index *index*.

When the dates for adjusted closing prices did not correspond to the components of the polarity vector with shifted dates, the correlation was defined as not available (NA). The analysis of data was performed on a computer with the following characteristics: 2.2 GHz Intel Core i7 processor,

**Table 8** Example of polarity vector with a date shift = - 7

Date	S140	Bing +	Bing-	NRC +	NRC-	Affin	SenticNet	Adj.Close
2020/May/12	0.000	0.29	0.25	0.48	0.30	- 0.02	- 0.58	2891.56
2020/May/13	0.000	0.23	0.33	0.51	0.38	- 0.03	0.59	2898.05

8 MB RAM, Mac OS 10.15.4. The method was based on the Python programming language.

## Results

Due to the large number of computed shifted correlations between the daily polarity of Twitter accounts and the daily behaviour of financial indices, Table 8 shows an example of the polarity vectors (*Date*, *S140*, *Bing*<sup>+</sup>, *Bing*<sup>-</sup>, *NRC*<sup>+</sup>, *NRC*<sup>-</sup>, *Affin*, *SenticNet*). These polarity vectors were found in the Bloomberg data-set, i.e., the @business Twitter account. Neutral polarity was determined when the S140 lexicon was used. Almost the same positive and negative polarity strengths were found when Bing Liu's lexicon was used, whereas with the NRC lexicon, the positive polarity strength was found to be greater than negative polarity strength. With the AFFIN Sentiment Lexicon, a slightly negative polarity was computed. the SenticNet lexicon detects the highest polarity. The date shift was equal to - 7. This means that the original dates of publications of the tweets were May 5, 2020, and May 6, 2020, whereas the values of the adjusted closing prices occurred on May 12, 2020, and May 13, 2020, respectively. The adjusted closing prices corresponded to Shanghai, China's financial index.

An example of the calculated correlations between the polarity and the adjusting closing price  $r(Mpricesentic)$  is shown in Table 9. This example shows a perfect correlation between the polarities and financial indices found in the Bloomberg data set.

After calculating the correlations of each financial indicator with the polarity generated by the proposed method  $r(Mpricesentic)$ , we identified the best correlations to understand financial market behaviour and sentiment analysis (Table 10). With SenticNet, we always found correlations between the polarity on Twitter and the behaviour of financial indices. However, there were some days when we did not find correlations with the lexicon-based analysis; for this reason, we decided to extend the search interval.

Despite extending the number of days of the search interval to find correlations with lexicon-based analysis, the SenticNet lexicon always generated the best results.

## Discussion

Sentic computing has the potential to be an important tool to improve decision-making in finance because it extracts and identifies patterns on Twitter [1, 15]. The analysis of polarity on Twitter has the potential to predict the behaviour of the financial market [16]. 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.

One of the applications of understanding the emotions of social media users consists of predicting the behaviour of financial indices. This reduces risk and uncertainty for investors. Investment portfolios can be formed with shares that follow the tendency of the indices, diversifying the selection of shares of different world markets. This can lead to an increase in the availability of more accurate and efficient sentiment measures and more wide-ranging studies.

Some of the analysed Twitter accounts were unrelated to financial indices due to several factors: (1) they had few followers; (2) they published very few posts or did not publish at all; (3) the Twitter posts were very specific and could affect only one sector and therefore were not perceived in the financial indices; and (4) the Twitter account did not publish relevant information for financial markets. For these reasons, some Twitter accounts were not included in the correlation analysis.

High correlations were found for the Investing.com, Bloomberg and CNN Business accounts. One of the reasons they had high correlations was their considerable number of followers—Investing.com, 168,000; CNN Business, 1.8 million; and Bloomberg, 6.4 million—and an average of 10 posts per day. Another Twitter account with high correlations was

**Table 9** Correlations found for the data set Bloomberg, date shift = - 7

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

**Table 10** Most significant correlations in absolute value for each data set

Data set	H1N1 (2009)			COVID-19 (2020)		
	Shift	$\max_{\text{shift-date}}  r $	lexicon	Shift	$\max_{\text{shift-date}}  r $	lexicon
CNNBusiness	− 2	0.46	SenticNet	− 6	0.70	SenticNet
Bloomberg	+2	0.54	SenticNet	− 7	1.00	SenticNet
NYTimes	− 1	0.39	SenticNet	− 13	0.97	SenticNet
Investing.com	NA	NA	SenticNet	4	0.93	SenticNet

The New York Times (46.8 million followers), covering both financial and economic topics, such as music, culture, sports, art and entertainment. Different types of content led to different kinds of sentiments of investors.

We found that in 2009, information posted on Twitter about H1N1 was practically non-existent; this was not the case with COVID-19 in 2020. Therefore, our findings show that some Twitter posts had a more significant effect in the COVID-19 era than in the H1N1 period. However, the effect took fewer days than the moderate effect in 2009. This was because the use of Twitter was more widespread in 2020, and more accounts published information regarding COVID-19 and its effects on finances. Investors had more information for decision-making.

The sentiment of investors could also predict the evolution of indices a few days in advance. This investigation revealed that it took 0 to 10 days for markets to react to information shared and disseminated on Twitter during the COVID-19 pandemic. During the H1N1 pandemic, this period was from 0 to 15 days [52]. We found correlations not only in the positive shift values but also in the negative ones (from − 11 to − 1). This indicates that the behaviour of the stock market affects the reactions of Twitter users. In the case of H1N1, this took 1 to 11 days, and in the case of COVID-19, this took 1 to 6 days (see Appendix).

Some studies reflect the importance of investors' emotions in financial markets [15]. In this research, we confirmed the argument of Ahmed [34] concerning investors' reactions to a financial event in terms of stock price movements. When applied to Twitter, days later, there is an effect on the reactions generated by Twitter posts on the topic of financial indices. 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 have been much greater than those of H1N1.

We confirm that the number of followers on Twitter influences the performance of financial indices. Regardless of whether the publication is accurate or not, this alters the sentiment of a higher number of investors, forcing them to seek a risk-reducing strategy. We verified that The New York Times is an important source for accurately predicting the behaviour of the stock market. Moreover, we found three more sources: CNN News, Bloomberg and

Investing.com. We also found higher shifted correlations with the SenticNet lexicon, and we detected that this method also produces a large number of cases in which correlations can be observed.

## Conclusions

This research aimed to study the effect of polarity in Twitter posts on the behaviour of world financial indices during pandemics through a lexicon-based approach. Our research question was as follows: How does the polarity generated by Twitter posts influence the behaviour of financial indices during pandemics? Our analysis shows sufficient evidence supporting the notion that Twitter posts have influenced financial indices during both pandemics, with remarkable influences in the case of COVID-19. We found that the effect of social media publications was more significant during the COVID-19 pandemic than during the H1N1 pandemic.

Our contributions are as follows. First, we found that the combination of a lexicon-based approach is enhanced by a shifted correlation analysis, as latent or hidden correlations can be found in data. Second, there is an important effect of sentiments on Twitter on financial indices, and this effect is observed a few days after the information is posted on Twitter. Third, our research confirms that SenticNet performed better than other lexicons. Fourth, we proposed a correlation matrix that contains the polarity of Twitter posts and the behaviour of financial markets to study their relationship when the date is shifted in terms of financial sentiment analysis and the adjusted closing prices. Fifth, our data show some important effects days after and days before the posts were made. Sixth, we also confirm that social media propagation—more Twitter accounts—over this period of 11 years has a direct impact on the indices' behaviour. Seventh, we presented the state of the art on sentic computing and finance.

Our findings confirm that the drop in stock prices during the COVID-19 era, compared with the H1N1 period, was more dramatic because there was more speculation, rumours and negative news. This investigation has additional value because it was conducted in the non-normal context of pandemics; similar research has been conducted

under normal or traditional conditions. This research can be extended by analysing other indices, markets, or products, such as cryptosystems, FOREX and futures; moreover, further analysis of investor reactions is needed. Such research can use of sophisticated techniques such as sentic computing and explore different kinds of emotions. Finally, it is possible to analyse data from other social

media platforms such as Facebook and WhatsApp in real time.

## Appendix

See Table 11, Table 12, Table 13, Table 14, Table 15, Table 16 and Table 17.

**Table 11** Correlations/date shifts for data set (CNNBusiness 2009)

Financial indices	S140	Bing +	Bing-	Nrc +	Nrc-	Affin	SenticNet
IPC	0.034/0	NA	NA	NA	NA	0.188/− 2	− 0.527/0
NASDAQ IXIC	− 0.048/− 2	NA	NA	NA	NA	0.198/− 2	− 0.489/0
DOW JONES DJI	0.126/0	NA	NA	NA	NA	0.339/− 2	− 0.68/0
BOVESPA	0.080/− 2	NA	NA	NA	NA	0.334/− 2	− 0.493/− 7
S&P GSPC	0.061/− 1	NA	NA	NA	NA	0.322/− 2	− 0.453/0
CAC 40 FRANCIA FCHI	0.199/0	NA	NA	NA	NA	0.405/− 2	0.597/10
DAX ALEMANIA GDAXI	0.183/− 3	NA	NA	NA	NA	0.359/− 2	0.563/10
HONG KONG HSI	− 0.111/− 3	NA	NA	NA	NA	0.263/− 2	− 0.449/− 11
NIKKEI JAPAN N225	− 0.081/0	NA	NA	NA	NA	0.469/− 2	0.587/− 6
SHANGHAI CHINA	− 0.095/− 1	NA	NA	NA	NA	− 0.413/− 2	− 0.573/− 4

**Table 12** Correlations/date shifts for data set (CNNBusiness 2020)

Financial indices	S140	Bing +	Bing-	Nrc +	Nrc-	Afinn	SenticNet
IPC	0.122/− 6	− 0.190/− 6	0.046/− 6	− 0.264/− 6	0.011/− 6	− 0.453/− 6	1.000/7
NASDAQ IXIC	− 0.073/− 6	− 0.594/− 6	− 0.003/− 6	− 0.401/− 6	− 0.086/− 6	− 0.621/− 6	− 1.000/8
DOW JONES DJI	− 0.068/− 6	− 0.694/− 6	− 0.001/− 6	− 0.282/− 6	− 0.059/− 6	− 0.594/− 6	− 1.000/8
BOVESPA	− 0.016/− 6	− 0.682/− 6	0.004/− 6	− 0.219/− 6	0.018/− 6	− 0.568/− 6	1.000/10
S&P GSPC	− 0.071/− 6	− 0.659/− 6	0.007/− 6	− 0.326/− 6	− 0.053/− 6	− 0.622/− 6	1.000/8
CAC 40 FRANCIA FCHI	− 0.003/− 6	− 0.621/− 6	− 0.196/− 6	− 0.205/− 6	− 0.267/− 6	− 0.351/− 6	− 1.000/11
DAX ALEMANIA GDAXI	− 0.041/− 6	− 0.605/− 6	− 0.068/− 6	− 0.214/− 6	− 0.140/− 6	− 0.479/− 6	− 1.000/8
HONG KONG HSI	− 0.128/− 6	− 0.704/− 6	− 0.054/− 6	− 0.272/− 6	− 0.085/− 6	− 0.530/− 6	− 1.000/8
NIKKEI JAPAN N225	− 0.111/− 6	− 0.553/− 6	− 0.226/− 6	− 0.395/− 6	− 0.360/− 6	− 0.317/− 6	− 1.000/6
SHANGHAI CHINA	− 0.183/− 6	− 0.589/− 6	0.060/− 6	− 0.405/− 6	− 0.074/− 6	− 0.668/− 6	− 1.000/6

**Table 13** Correlations/date shifts for data set (Bloomberg 2009)

Financial indices	S140	Bing +	Bing-	Nrc +	Nrc-	Afinn	SenticNet
IPC	0.391/2	0.050/7	0.018/7	0.050/0	0.018/0	0.309/2	− 0.500/3
NASDAQ IXIC	0.479/2	0.425/2	0.012/2	0.447/8	0.012/8	0.345/− 1	− 0.697/3
DOW JONES DJI	0.493/0	0.458/2	0.156/8	0.471/8	0.156/2	0.168/0	− 0.522/3
BOVESPA	NA	NA	NA	NA	NA	NA	− 0.497/2
S&P GSPC	0.540/6	0.429/8	0.112/8	0.444/6	0.112/6	0.205/5	− 0.530/3
CAC 40 FRANCIA FCHI	0.488/1	0.068/1	0.262/1	0.057/2	0.262/1	− 0.005/0	− 0.502/15
DAX ALEMANIA GDAXI	0.476/0	0.027/11	0.199/11	0.018/2	0.199/8	0.094/8	0.543/15
HONG KONG HSI	0.379/2	0.190/0	0.087/2	0.189/2	0.087/2	0.306/2	0.589/12
NIKKEI JAPAN N225	0.313/0	0.113/6	0.036/6	0.114/2	0.036/6	0.008/5	0.435/12
SHANGHAI CHINA	− 0.139/− 1	0.183/− 1	− 0.279/− 1	0.199/− 6	− 0.279/6	0.484/6	− 0.493/5



**Table 14** Correlations/date shifts for data set (Bloomberg 2020)

Financial indices	S140	Bing +	Bing-	Nrc +	Nrc-	Affin	SenticNet
IPC	NA	NA	NA	NA	NA	NA	1.000/7
NASDAQ IXIC	NA	NA	NA	NA	NA	NA	– 1.000/8
DOW JONES DJI	NA	NA	NA	NA	NA	NA	– 1.000/8
BOVESPA	NA	NA	NA	NA	NA	NA	1.000/– 1
S&P GSPC	NA	NA	NA	NA	NA	NA	– 1.000/10
CAC 40 FRANCIA FCHI	NA	NA	NA	NA	NA	NA	– 1.000/10
DAX ALEMANIA GDAXI	NA	NA	NA	NA	NA	NA	– 1.000/11
HONG KONG HSI	NA	NA	NA	NA	NA	NA	– 1.000/6
NIKKEI JAPAN N225	NA	NA	NA	NA	NA	1.000/– 7	– 1.000/6
SHANGHAI CHINA	NA	NA	NA	NA	NA	1.000/– 7	– 1.000/6

**Table 15** Correlations/date shifts for data set (NYTimes 2009)

Financial indices	S140	Bing +	Bing-	Nrc +	Nrc-	Affin	SenticNet
IPC	0.349/0	NA	NA	NA	NA	– 0.181/0	– 0.527/0
NASDAQ IXIC	0.220/0	NA	NA	NA	NA	– 0.188/0	0.489/0
DOW JONES DJI	0.391/3	NA	NA	NA	NA	– 0.174/3	– 0.468/3
BOVESPA	0.346/– 7	NA	NA	NA	NA	– 0.057/– 7	– 0.493/– 7
S&P GSPC	0.340/0	NA	NA	NA	NA	– 0.177/0	– 0.453/0
CAC 40 FRANCIA FCHI	0.356/10	NA	NA	NA	NA	– 0.130/10	0.598/10
DAX ALEMANIA GDAXI	0.398/10	NA	NA	NA	NA	– 0.157/10	0.563/10
HONG KONG HSI	0.206/– 11	NA	NA	NA	NA	– 0.037/– 11	– 0.449/– 11
NIKKEI JAPAN N225	0.086/– 6	NA	NA	NA	NA	0.116/– 6	0.588/– 6
SHANGHAI CHINA	– 0.297/– 4	NA	NA	NA	NA	– 0.091/– 4	– 0.573/– 4

**Table 16** Correlations/date shifts for data set (NYTimes 2020)

Financial indices	S140	Bing +	Bing-	Nrc +	Nrc-	Affin	SenticNet
IPC	0.637/– 13	NA	NA	NA	NA	– 0.079/– 13	– 1.000/10
NASDAQ IXIC	0.631/– 13	NA	NA	NA	NA	– 0.299/– 13	– 1.000/10
DOW JONES DJI	0.849/– 13	NA	NA	NA	NA	– 0.036/– 13	– 1.000/10
BOVESPA	0.724/– 13	NA	NA	NA	NA	– 0.695/– 13	– 1.000/11
S&P GSPC	0.766/– 13	NA	NA	NA	NA	0.123/– 13	– 1.000/10
CAC 40 FRANCIA FCHI	0.949/– 13	NA	NA	NA	NA	– 0.246/– 13	– 1.000/11
DAX ALEMANIA GDAXI	0.973/– 13	NA	NA	NA	NA	– 0.391/– 13	– 1.000/11
HONG KONG HSI	0.113/– 13	NA	NA	NA	NA	0.422/– 13	1.000/10
NIKKEI JAPAN N225	0.271/– 13	NA	NA	NA	NA	0.037/– 13	– 1.000/8
SHANGHAI CHINA	0.635/– 13	NA	NA	NA	NA	– 0.119/– 13	– 1.000/8

**Table 17** Correlations/date shifts for data set (Investing.com 2020), date shift = 4

Financial indices	S140	Bing +	Bing-	Nrc +	Nrc-	Affin	SenticNet
IPC	− 0.181	NA	NA	NA	NA	− 0.875	NA
NASDAQ IXIC	− 0.052	NA	NA	NA	NA	− 0.825	NA
DOW JONES DJI	0.918	NA	NA	NA	NA	0.247	NA
BOVESPA	0.752	NA	NA	NA	NA	0.710	NA
S&P GSPC	0.545	NA	NA	NA	NA	− 0.374	NA
CAC 40 FRANCIA FCHI	0.461	NA	NA	NA	NA	0.628	NA
DAX ALEMANIA GDAXI	0.694	NA	NA	NA	NA	− 0.356	NA
HONG KONG HSI	0.555	NA	NA	NA	NA	0.585	NA
NIKKEI JAPAN N225	− 0.136	NA	NA	NA	NA	− 0.556	NA
SHANGHAI CHINA	− 0.280	NA	NA	NA	NA	− 0.938	NA

## Compliance with Ethical Standards

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

## References

- Cambria E, Hussain A. Sentic computing: a common-sense-based framework for concept-level sentiment analysis. Sentic Computing. Switzerland: Springer; 2015.
- Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *J Big Data*. 2015;2(1):1.
- Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M. Sentiment embeddings with applications to sentiment analysis. *IEEE Trans Knowl Data Eng*. 2015;28(2):496–509.
- Qian B, Rasheed K. Stock market prediction with multiple classifiers. *Appl Intell*. 2007;26(1):25–33.
- Attigeri G V, MM MP, Pai RM, Nayak A. Stock market prediction: a big data approach. In: TENCON 2015–2015 IEEE Region 10 Conference. 2015. p. 1–5.
- Khatua A, Khatua A, Cambria E. A tale of two epidemics: contextual Word2Vec for classifying twitter streams during outbreaks. *Inf Process Manag*. 2019;56(1):247–57.
- Shiller RJ. Measuring bubble expectations and investor confidence. *J Psychol Financ Mark*. 2000;1(1):49–60.
- Kurov A. Investor sentiment, trading behavior and informational efficiency in index futures markets. *Financ Rev*. 2008;43(1):107–27.
- Oberlechner T, Hocking S. Information sources, news, and rumors in financial markets: insights into the foreign exchange market. *J Econ Psychol*. 2004;25(3):407–24.
- Cruz FMD, Gomes MYFSDF. The influence of rumors in the stock market: a case study with Petrobras. *Transinformaçao*. 2013;25(3):187–93.
- Doerr B, Fouz M, Friedrich T. Why rumors spread so quickly in social networks. *Commun ACM*. 2012;55(6):70–5.
- Lipsitch M, Finelli L, Heffernan RT, Leung GM, for the 2009 H1N1 Surveillance Group SC. Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1. *Biosecurity bioterrorism biodefense Strateg Pract Sci*. 2011;9(2):89–115.
- Zhang D, Hu M, Ji Q. Financial markets under the global pandemic of COVID-19. *Financ Res Lett*. 2020;101528.
- Hassan TA, Hollander S, van Lent L, Tahoun A. Firm-level exposure to epidemic diseases: Covid-19, SARS, and H1N1. 2020.
- Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: from unimodal analysis to multimodal fusion. *Inf Fusion*. 2017;37:98–125.
- Cambria E, Das D, Bandyopadhyay S, Feraco A. Affective computing and sentiment analysis. In: A practical guide to sentiment analysis. Springer; 2017. p. 1–10.
- Li G. Information quality, learning, and stock market returns. *J Financ Quant Anal*. 2005;595–620.
- Kearney C, Liu S. Textual sentiment in finance: a survey of methods and models. *Int Rev Financ Anal*. 2014;33:171–85.
- Jiao P, Veiga A, Walther A. Social media, news media and the stock market. *J Econ Behav Organ*. 2020;176:63–90.
- Bandelj N. Emotions in economic action and interaction. *Theory Soc*. 2009;38(4):347–66.
- Martínez-Miranda J, Aldea A. Emotions in human and artificial intelligence. *Comput Human Behav*. 2005;21(2):323–41.
- Atzmueller M. Mining social media: key players, sentiments, and communities. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2012;2(5):411–9.
- Nguyen TH, Shirai K, Velcin J. Sentiment analysis on social media for stock movement prediction. *Expert Syst Appl*. 2015;42(24):9603–11.
- Poria S, Cambria E, Howard N, Huang G Bin, Hussain A. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*. 2016;174:50–9. <https://doi.org/10.1016/j.neucom.2015.01.095>.
- Cambria E, Fu J, Bisio F, Poria S. Affectivespace 2: Enabling affective intuition for concept-level sentiment analysis. In: AAAI. 2015. p. 508–14.
- Liu B. Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol*. 2012;5(1):1–167.
- Susanto Y, Livingstone AG, Ng BC, Cambria E. The hourglass model revisited. *IEEE Intell Syst*. 2020;35(5):96–102.
- Kraus M, Feuerriegel S. Sentiment analysis based on rhetorical structure theory: learning deep neural networks from discourse trees. *Expert Syst Appl*. 2019;118:65–79.
- 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. p. 185–209.
- Cambria E, Poria S, Gelbukh A, Thelwall M. Sentiment analysis is a big suitcase. *IEEE Intell Syst*. 2017;32(6):74–80.
- Calvo RA, D'Mello S, Gratch JM, Kappas A. The Oxford handbook of affective computing. USA: Oxford University Press; 2015.
- Cambria E, Hussain A, Havasi C, Eckl C. AffectiveSpace : blending common sense and affective knowledge to perform emotive reasoning. *WOMSA 1st Work Opin Min Sentim Anal Opin Min Sentim Anal Work Opin Min*. 2009:32–41.

33. 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). 2016. p. 1127–34.
34. Ahmed SM. Quantification of investor emotion in financial news by analyzing the stock price reaction. 2017 Int Conf Inf Commun Technol ICICT 2017. 2017. p. 119–23.
35. Minhas S, Poria S, Hussain A, Hussainey K. A review of artificial intelligence and biologically inspired computational approaches to solving issues in narrative financial disclosure. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2013;7888 LNAI:317–27.
36. Atzeni M, Dridi A, Reforgiato Recupero D. Using frame-based resources for sentiment analysis within the financial domain. *Prog Artif Intell*. 2018;7(4):273–94. <https://doi.org/10.1007/s13748-018-0162-8>.
37. Xing FZ, Cambria E, Welsch RE. Intelligent asset allocation via market sentiment views. *IEEE Comput Intell Mag*. 2018a;13(4):25–34.
38. Picasso A, Merello S, Ma Y, Oneto L, Cambria E. Technical analysis and sentiment embeddings for market trend prediction. *Expert Syst Appl*. 2019;135:60–70. <https://doi.org/10.1016/j.eswa.2019.06.014>.
39. Xing FZ, Cambria E, Welsch RE. Natural language based financial forecasting: a survey. *Artif Intell Rev*. 2018b;50(1):49–73.
40. Malandri L, Xing FZ, Orsenigo C, Vercellis C, Cambria E. Public mood-driven asset allocation: the importance of financial sentiment in portfolio management. *Cognit Comput*. 2018;10(6):1167–76.
41. Upreti BR, Back PM, Malo P, Ahlgren O, Sinha A. Knowledge-driven approaches for financial news analytics. In: *Network theory and agent-based modeling in economics and finance*. Springer; 2019. p. 375–404.
42. Dridi A, Atzeni M, Reforgiato Recupero D. FineNews: fine-grained semantic sentiment analysis on financial microblogs and news. *Int J Mach Learn Cybern*. 2019;10(8):2199–207. <https://doi.org/10.1007/s13042-018-0805-x>.
43. Merello S, Ratto AP, Oneto L, Cambria E. Ensemble application of transfer learning and sample weighting for stock market prediction. *Proc Int Jt Conf Neural Networks*. 2019:1–8.
44. Xing FZ, Cambria E, Welsch RE. Growing semantic vines for robust asset allocation. *Knowledge-Based Syst*. 2019;165:297–305. <https://doi.org/10.1016/j.knsys.2018.11.035>.
45. Xing FZ, Cambria E, Zhang Y. Sentiment-aware volatility forecasting. *Knowledge-Based Syst*. 2019;176:68–76. <https://doi.org/10.1016/j.knsys.2019.03.029>.
46. Akhtar MS, Ekbal A, Cambria E. How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble. *IEEE Comput Intell Mag*. 2020;15(1):64–75.
47. Smimou K, Bector CR, Jacoby G. Portfolio selection subject to experts' judgments. *Int Rev Financ Anal*. 2008;17(5):1036–54.
48. Kiritchenko S, Zhu X, Mohammad SM. Sentiment analysis of short informal texts. *J Artif Intell Res*. 2014;50:723–62.
49. Mohammad SM, Turney PD. Nrc emotion lexicon. *Natl Res Counc Canada*. 2013;2.
50. Keshavarz H, Abadeh MS. ALGA: adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs. *Knowledge-Based Syst*. 2017;122:1–16.
51. Cambria E, Li Y, Xing FZ, Poria S, Kwok K. SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020. p. 105–14.
52. Kiyamaz H. The effects of stock market rumors on stock prices: evidence from an emerging market. *J Multinatl Financ Manag*. 2001;11(1):105–15.