Journal of Business Economics and Management
ISSN 1611-1699 / eISSN 2029-4433
2023 Volume 24 Issue 1: 177–198
https://doi.org/10.3846/jbem.2023.18805

CRITICAL REVIEW OF TEXT MINING AND SENTIMENT ANALYSIS FOR STOCK MARKET PREDICTION

Zuzana JANKOVÁ*

Institute of Informatics, Faculty of Business and Management, Brno University of Technology, Brno, Czech Republic

Received 04 July 2022; accepted 09 March 2023

Abstract. The paper is aimed at a critical review of the literature dealing with text mining and sentiment analysis for stock market prediction. The aim of this work is to create a critical review of the literature, especially with regard to the latest findings of research articles in the selected topic strictly focused on stock markets represented by stock indices or stock titles. This requires examining and critically analyzing the methods used in the analysis of sentiment from textual data, with special regard to the possibility of generalization and transferability of research results. For this reason, an analytical approach is also used in working with the literature and a critical approach in its organization, especially for completeness, coherence, and consistency. Based on the selected criteria, 260 articles corresponding to the subject area are selected from the world databases of Web of Science and Scopus. These studies are graphically captured through bibliometric analysis. Subsequently, the selection of articles was narrowed to 49. The outputs are synthesized and the main findings and limits of the current state of research are highlighted with possible future directions of subsequent research.

Keywords: bibliometric analysis, financial market, literature review, sentiment analysis, stock market, text mining.

JEL Classification: C53, C63, G11.

Introduction

Due to the growing progress in the analysis of social media, it was possible to include online news as an input to models predicting future developments. In addition, the dynamic development of unstructured data processing techniques now allows more complex information to be extracted. Digitized text in the form of unstructured data is growing at a dizzying rate in terms of volume, availability and information contained in them. However, traditional approaches to qualitative text analysis are not able to process this volume within an acceptable time horizon and with available computational resources. For this reason, attention is focused

*Corresponding author. E-mail: *Zuzana.Jankova@vutbr.cz*

on text mining, which is a set of approaches to examine and analyze large collections of texts according to Antons et al. (2020). The use of social media to predict the stock market can be considered a relatively new and revolutionary approach. However, the enormous volumes of information that are published online every second can cause significant computational and technical problems, according to Bustos and Pomares-Quimbay (2020). In particular, sentiment analysis is currently considered to be one of the fastest growing areas of IT research, which makes it difficult to monitor all activities in this area. The beginnings of sentimental analysis date back to the early 20th century. However, only with the growing number of online texts, contributions, and news on the web did computer sentimental analysis develop. In recent years, the analysis of sentiment has shifted from the analysis of online product reviews to social media texts from Twitter and Facebook according to Mäntylä et al. (2018). Through the extraction of valuable information, a financial analyst can find a relationship between the information extracted and predict the behavior of financial markets and thus achieve excessive profits.

A large number of scientific and research works are currently available that examines the use of textual data mining and the performance of sentimental analysis with consequent influence and predictive power on financial markets. There are inconsistent outputs regarding the very effect of sentiment on stock markets. The study by Nguyen et al. (2015) report that sentiment extracted from social networks has little or no power to predict the performance of stock markets, while other studies such as Nti et al. (2020), on the contrary, claim that social media sentiment shows a relatively strong predictive ability. Shi et al. (2018) found that investor sentiment has short-term positive effects and medium-term reverse effects on the stock market. An asymmetric effect was also found, with high investor sentiment gaining a more pronounced effect on stock returns. According to the study by Hwang and Kim (2019), the development of stock prices affects sentiment, or the authors concluded that stock prices respond to social issues earlier than sentiment, which is subsequently expressed by market participants in text messages. Nguyen et al. (2015) point out that sentiment may not be a reason for price movements in the stock market. However, although sentiment may be one of the factors influencing price movements in the stock market, extracted sentiments from discussion forums, etc. do not reflect the price due to chaotic, erroneous commentary or predictions of human error when publishing news.

As mentioned above, the use of extracted sentiment from social media to predict prices in the stock market is still an open research issue. However, few studies have focused on reviewing the literature on this topic. The identified gap in current research points to the fact that most existing review papers insufficiently cover the broad area of sentiment on stock markets, especially in terms of orientation to a specific application in the financial market or orientation to a specific sentiment extraction algorithm. The aim of this work is to create a complex critical review of the literature, especially regarding the latest international findings of research articles in the selected topic, strictly focused on stock markets represented by stock indices or stock titles. This requires examining and critically analyzing the methods used in the analysis of sentiment from textual data, with special regard to the possibility of generalization and transferability of research results. For this reason, an approach of bibliometric analysis is also used in working with the literature and a critical approach in its organization, especially for completeness, coherence, and consistency. In particular, bibliometric analysis is chosen for the visual assessment of the most common keywords, authors and publications in the issue, which provides a basic insight into the issue and contributes to the originality of the presented review of current literature. Thanks to graphical processing, bibliometric analysis makes it possible to uncover hidden, not entirely obvious in the literature, links and implications across research studies. And it identifies authoritative researchers and authors who are subjected to a critical content analysis of the results of their researches. Last but not least, it is necessary to identify still open problems and also which areas need additional attention and re-examination in order to provide relevant and clear conclusions.

The layout of the article is as follows: Section 1 presents the methodology and approach of data source collection. Section 2 represents a bibliometric analysis of selected articles according to the selected software. Section 3 represents a critical review of the literature with a detailed description of the findings in the selected area. Section 4 discusses and synthesizes the main findings and limits of current scientific work in the field of text mining and financial market sentiment analysis.

1. Methodology and data collection

Due to the fact that there is a wide range of scientific articles and contributions dealing with various topics, it is highly important to select the appropriate knowledge base of information related to the issue. The following section describes the method of selection of relevant data sources and also the used methodology of bibliometric network processing, including the selection of a suitable visualization tool, which will serve as a tool for mapping key authors, sources, references and keywords related to the issue.

1.1. Data collection

To provide a structured overview of the state and development of the application of text mining and sentimental analysis in the stock markets, a systematic review of the literature is performed, which was performed as follows. As a first step in the systematic review, two databases of primary scientific papers, Web of Science and Scopus, were selected. As a second step, a list with commonly used keywords was created, which are use to denote the application of stock market prediction techniques using sentiment analysis and text extraction. Keywords includes terms that e.g., “text mining”; “sentiment analysis” and terms related to the stock market e.g., “stock”; “stock market”; “financial market”. Only articles selected on the basis of the list of keywords are selected for subsequent analysis.

The selection process is shown in Figure 1. After starting a database search, 260 articles matched the search equation, with 152 articles identified from the Scopus database and 108 articles from the Web of Science database. 70 duplicate articles and posts were excluded from this search. After reading the titles and their abstracts, a further 111 articles were excluded from this review because they are not focused to predict the direction of the stock market (69), they are not published in English (3) and some articles were unavailable (39). After reading these articles in full, 30 papers were rejected because they did not provide the necessary data. A critical review is performed on 49 scientific articles and papers that have been identified as relevant to the research.

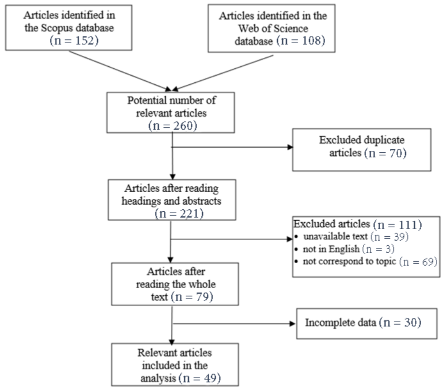

Figure 1. Diagram of relevant article selection processes

1.2. Research method

Bibliometric analysis is used to analyze large collections of scientific and academic works and is mainly used to reveal hidden or at first glance barely recognizable citation and co-citation relationships across publications. Bibliometric analysis integrates not only scientific, but especially statistical methods, which is able to generate bibliometric maps and visualize the appropriate links in this area. Data sources from various paid and unpaid world databases can be used to apply bibliometric analysis. The database of scientific articles plays a very important role in the field of scientific research. As a widely used bibliometric software VOSViewer, capable of displaying large bibliometric maps (van Eck & Waltman, 2010). Bibliometric maps can be a useful tool to find potential hidden links between scientific articles and thus provide a basis for scientific research.

Bibliometric maps display intuitively, and clearly bibliometric information of individual publications obtained from the Web of Science and Scopus databases. The importance of individual publications, their authors, keywords used or influential journals in which articles are published can be revealed by citation and co-citation analysis and graphically visualized through bibliometric mapping.

2. Bibliometric analysis

Bibliometric analysis is performed on relevant articles, the selection of which is given in Section 1 and processed by VOSViewer. This software visualizes bibliometric maps of large data sources and synthesizes them for clustering (van Eck & Waltman, 2010). Bibliometric analysis is applied to the analysis of keywords in Section 2.1, then to the analysis of co-citations of journals in Section 2.2 and finally to the co-citation of references in Section 2.3.

2.1. Visualization of keyword analysis

Keyword occurrence analysis analyzes the common occurrence of all keywords. For this purpose, the full counting method is used, in other words, it calculates each occurrence of the keyword in a particular document. To create the bibliometric method, only networks were selected where the keyword occurred 5 or more times. This narrowed down to 173 keywords, which are then visualized on the final bibliometric map.

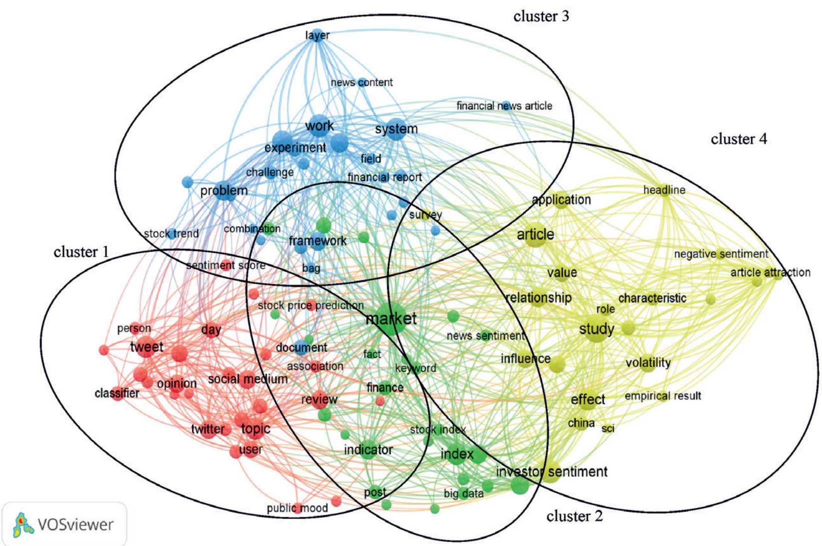

Figure 2. Results of bibliometric keyword occurrence analysis

The bibliometric map captures bubbles of different sizes and colors, which are interconnected. Color resolution separates the groups, which are referred to as clusters. Individual clusters identify highly linked keywords. The larger the bubble shown, the higher the frequency of use of the keyword and vice versa. Furthermore, the different lengths of linking both keywords and individual clusters can be seen. The shorter the drive distance, the stronger the association between keywords, including clusters, and vice versa.

Figure 2 shows the occurrence of keywords through bibliometric analysis. The total number of 173 keywords is divided into 4 color-coded clusters with a total of 700 lines. In the first cluster, terms such as "social media" are synthesized with references to "twitter", "tweet" and "opinion". This cluster contains the most keywords related to sentimental analysis and opinion analysis on social networks. The top keywords of the second cluster were "market", "index", "finance" and "indicator", i.e., terms related to the stock market. The third group includes "system", "experiment", "problem" and "stock trend". It indicates a problem with text data extraction and sentiment mining, which requires certain tools, systems, algorithms. Finally, the fourth cluster is the largest and includes the keyword used all "article", "study", "investor sentiment", i.e., it connects cluster 1 and cluster 2 and thus refers to the acquisition of data from text documents. The size of the bubble is greatest in the area of interest, which is the stock market, text data and investor sentiment. In particular, cluster 2 has the shortest links and has fairly close links to cluster 1, but also links to other clusters. Conversely, the greatest distance can be observed between clusters 1 and 4. Cluster 1 refers to social media such as Twitter, while cluster 4 refers more to traditional media such as articles and studies.

2.2. Visualization of co-citations journals

The analysis of co-citations of journals on sentimental analysis in stock markets is important because the frequency of co-citations between two sources shows the similarity and interconnectedness between the two sources that is given by their citations in the same document. The minimum number of co-citations was set at the level of two items. Each node expresses the source of the journal, and its size represents the number of citations. The interconnection between two nodes indicates a quote relationship. The nodes are differentiated by color, again it is true that magazines in the same cluster are more similar to each other.

Figure 3 shows that Journal of Finance has the highest number of co-citations with a total of 90. The scope of this journal includes studies in the field of stock market finance, so this result can be expected. Expert system with Application had the greatest total strength link 1503. Another high-strength journal (1269) is the Decision Support System. The following most cited journals were the Journal of Financial Economics, Review of Financial Studies, Neurocomputing, International Journal of Forecasting and Journal of Banking and Finance.

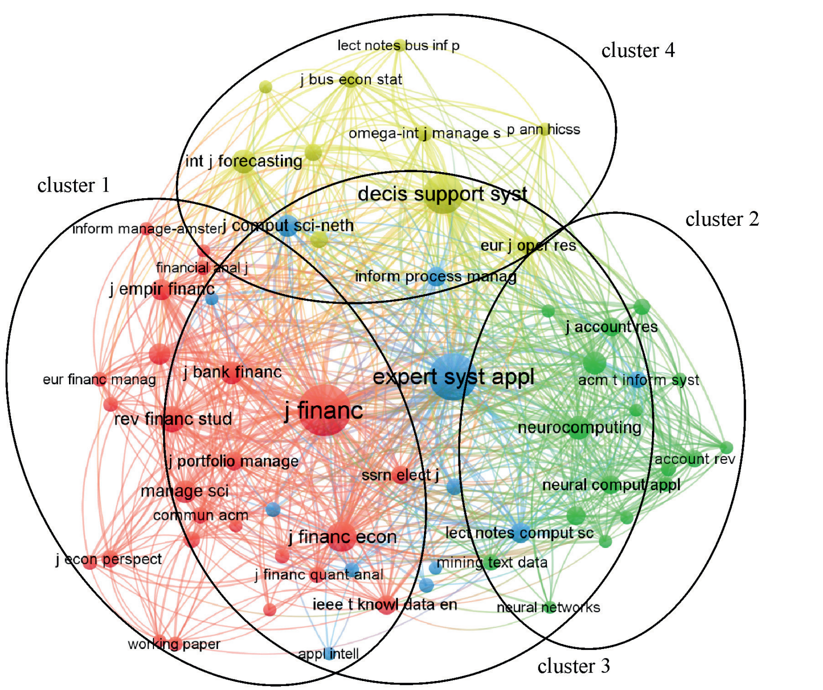

Figure 3. Results of bibliometric co-citations of journals analysis

VOS Viewer software classified the journals into four clusters. The red cluster contained 42 sources and included journals that deal mainly with finance, financial management and investing. The blue cluster basically focused on relevant research in the field of expert systems and learning algorithms and included 28 sources. The yellow cluster (16) was the smallest and contained mainly journals dealing with decision-making, prediction and operations research. Finally, the green cluster (26) differed most thematically from the others, as it consisted mainly of journals focusing on neuroscience, a leading expert system used in textual and sentimental analysis of the stock market.

2.3. Visualization of co-citations references

Another citation study analyzed the co-citations of references on sentimental analysis in stock markets. As Small (1973) describes the very frequently co-cited references contain basic and key ideas and findings in the field. For the correct visualization of the bibliometric map, the co-citation threshold was set to three. Figure 4 shows a total of 167 references that met the threshold. The size of a node represents the number of its citations. Citation basically indicates the frequency of citations between two cited articles. The VOS Viewer software used identified a total of four clusters. The most numerous red cluster contains a total of 51 references, the blue cluster 36 references. Both clusters contain key references in financial theory, specifically articles by Markowitz or Fama, which contain basic and fundamental starting points for the theory of financial markets. The yellow cluster contains 30 references, and the green cluster contains 46 cited references.

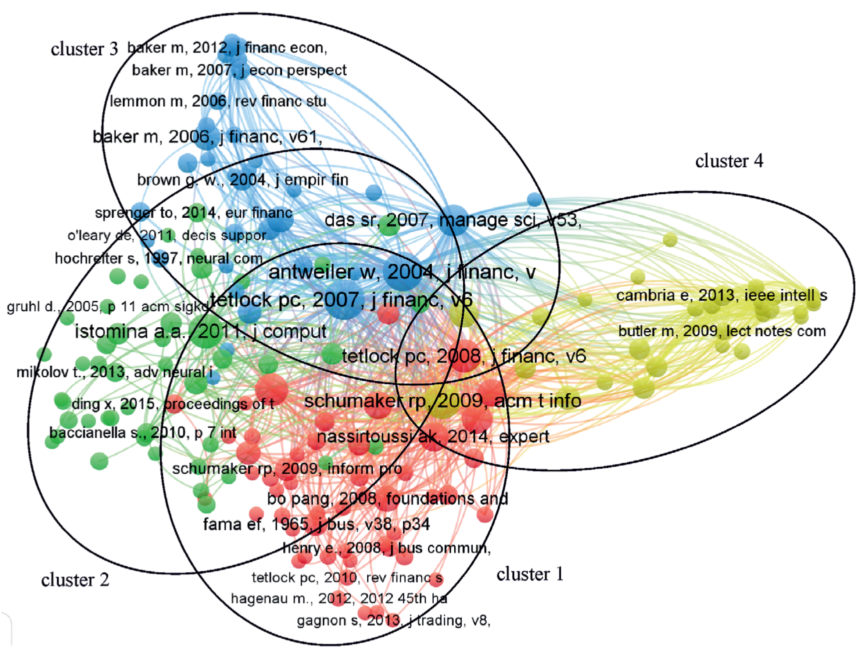

Figure 4. Results of bibliometric co-citations of references analysis

3. Critical review of the current state of the literature

The bibliometric analysis is followed by a critical review of the current state of knowledge of the selected articles. Studies dealing with stock market prediction can be divided into structured and unstructured according to the types of input variables. The vast majority of research is focused exclusively on structured data, which are simply processed by known and studied techniques. However, it is now possible to use unstructured data which contain additional information. In addition to focusing on input data represented in the form of a text source in Section 3.1, numerical data originating from the market are used as input variables for stock market prediction in Section 3.2. Attention is also focused on the most common types of sentiment classification algorithms, which are currently the most used in Section 3.3.

3.1. Sources of text unstructured data

Text data can have several sources and types of content, as shown in Table 1. Some of the studies used (25%) use major financial sites such as The Wall Street Journal as text data, Reuters, Dow Jones as well as Yahoo! Finance, Google Finance, NYSE and NASDAQ.

Table 1. Sources of text unstructured data (source: Janková, 2021)

Author	Type of corpus	Sources of corpus	No. of items
Hao et al. (2021)	Financial news	NYSE	2474
Owen and Oktariani (2020)	Social networks	StockTwits	1454
Sakhare et al. (2020)	Social networks	Twitter	n/a
Li et al. (2020)	Financial news	Eastmoney	18 million
Sun et al. (2020)	Microblog	Sina Weibo	22 504
Nti et al. (2020)	Social networks	Twitter, Google trends	thousands
Bouktif et al. (2020)	Social networks	Twitter	170 000/stock
Bouktif et al. (2019)	Social networks	Twitter	n/a
Birbeck and Cliff (2019)	Social networks	Twitter	1 474 747
Moro et al. (2019)	Social networks	Twitter	10 million
Jammalamadaka et al. (2019)	Financial news	Yahoo Finance, Google, …	n/a
Gross-Klussmann et al. (2019)	Social networks	Twitter	102 million
Derakhshan and Beigy (2019)	Financial news	Yahoo Finance, Sahamyab	787 547
Chen and Chen (2019)	Financial news	ChinaTimes, cnYES, ...	n/a
Ren et al. (2019)	Microblog	Sina, Eastmoney	1 930 592
Hwang and Kim (2019)	Financial news	CNBC, Reuers, Wall Stree	n/a
Chen and Shih (2019)	Financial news	Wantgoo	1312
Batra a Daudpota (2018)	Social networks	StockTwits	300 000
Eliacik and Erdogan (2018)	Social networks	Twitter	1 241 234
Kim et al. (2018)	Financial news	8-K	n/a
Feuerriegel and Gordon (2018)	Financial news	DGAP, EQS Group	80 813
Long et al. (2018)	Financial news	Eastmoney	5990

End of Table 1

Author	Type of corpus	Sources of corpus	No. of items
Hajek (2018)	Financial news	10-K	n/a
Shi et al. (2018)	Financial news	Eastmoney	5 163 210
Pagolu et al. (2017)	Social networks	Twitter	250 000
Ab. Rahman et al. (2017)	Financial news	Edge Markets	14 992
Khedr et al. (2017)	Financial news	NASDAQ, Reuters, …	n/a
Domeniconi et al. (2017)	Social networks	Twitter	10 million
Kraus and Feuerriegel (2017)	Financial news	8-K	34 782
Xie and Jiang (2017)	Financial news	CNBC, Reuers, Wall Street	2 302 692
Oliveira et al. (2017)	Social networks	Twitter	31 million
Urolagin (2017)	Social networks	Twitter	n/a
Simoes et al. (2017)	Social networks	Twitter	n/a
Alostad and Davulcu (2017)	Financial news	NASDAQ, Twitter	53,641+780,139
Zhao et al. (2016)	Microblog	Sina Weibo	6.1 million
Das and Das (2016)	Review	TripAdvisor	15 763
Eliacik and Erdogan (2016)	Social networks	Twitter	1 148 181
Hajek and Bohacova (2016)	Reports	Bank news	n/a
Al-Ramahi et al. (2015)	Financial news	Yahoo Finance	200
Nguyen et al. (2015)	Financial news	Yahoo Finance	n/a
Al Nasseri et al. (2015)	Social networks	StockTwits	n/a
Smailović et al. (2014)	Social networks	Twitter	152 570
Al Nasseri et al. (2014)	Social networks	StockTwits	2892
Meesad and Li (2014)	Social networks	Twitter	4622
Tirea and Negru (2013)	Financial news	M.bursa, M.antena3.	6217
Oliveira et al. (2013)	Social networks	Twitter	n/a
Nann et al. (2013)	Social networks	Twitter, Yahoo Finance	2 971 381
Siering (2012)	Financial news	Dow Jones News	11 518
O'Hare et al. (2009)	Financial news	n/a	232

The vast majority of these studies use financial reports rather than general reports, as they believe they contain less noise. These studies extract the text of the message itself or just its title. For example, as reported by Huang et al. (2010), news headlines are traditionally used. In addition to the leading financial websites, general financial reports or financial reports of specific companies were surveyed, such as Feuerriegel and Gordon (2018). Another section of researchers examined fewer formal sources of textual information, such as Li et al. (2020), Ren et al. (2019), Long et al. (2018), in their work, they looked at the text published and discussed at Eastmoney, which is the largest news exchange in China. It is a very popular Chinese economic website, which has the largest number of visitors of all financially oriented websites in China. Recently, however, they have been examining textual content from social media. This text data source makes up 39% of the analyzed sources. One group of researchers focuses only on Twitter and uses it to predict the market and analyze public sentiment more effectively (Sakhare et al., 2020). Gross-Klussmann et al. (2019) state that the directional sentimental measures on Twitter significantly reflect the current return trends of the stock market index. Expert users in particular are a major driver of the interdependence of Twitter sentiment and financial markets. Other authors have focused on the StockTwits platform making Stocktwits the largest social network for investors and traders with more than five million community members and millions of monthly visitors. As the key voice of "social financing", Stocktwits is the best way to determine what is going on in the markets and events you care about and transform the financial media for the next generation of investors. Al Nasseri et al. (2015) confirm that StockTwits contributions contain valuable information and prevent capital market.

Other researchers and researchers focused on discussion and stock forums, microblogs, online reviews, etc. This group makes up 14% of the analyzed text data from relevant studies. This is a minority group, which has been losing its attractiveness, especially in recent years, and is moving into the background of current research. For example, Sun et al. (2020); Zhao et al. (2016) who consider Sina Weibo to be a suitable source of financial reports. Sina Weibo is an alternative to the social network Twitter in China, as in 2009 this social platform was blocked in China.

A total of 22% of the authors used a combination of financial news and social media, or discussion forums. Nti et al. (2020) found that the combined dataset shows the highest accuracy, followed by the social network Twitter, web messages, then Google trends, and the lowest accuracy is with the forum post.

In Table 1, in addition to the type and source of the text corpus, the number of items of the examined text source is also given. It can be observed that this number ranges from hundreds to several million text source items. However, some sources do not indicate the size of the corpus they examined in their work. Alostad and Davulcu (2017) report that a higher volume of Tweets leads to a statistically significant increase in accuracy. However, Kraus and Feuerriegel (2017) point out common obstacles stemming from the fact that large test datasets are not commonly freely available. However, the incorporation of these large corpora is essential for creating accurate models. Due to the restriction of access to extensive data sources, the long-term influence of social networks for financial decision-making is still inconclusive.

3.2. Sources of numerical structured data

In addition to textual data, numerical data representing the price volatility or trading volume of a stock or stock indices are also used to predict the movements of stock market. The purpose of this data is to predict through selected algorithms. Table 2 provides key details on these market data. Research focused on prediction of stock market movement is represented either in the form of a stock market index (39%), individual stocks (8%), a group of indices or stocks (29%). It is striking that in a total of 24% of the relevant works, the name of the stock or index that was examined was not specifically mentioned.

In terms of stock indices, the US Dow Jones Industrial Average and the S&P 500 are often examined. In addition to US stock indices, Chinese indices such as CSI 300 have also been extensively studied as a proxy for the entire Chinese stock market. addition to this stock index, the SSE 50 Index and the Shanghai Composite Index are also examined in China. Minority studies also include studies focused on the Taiwanese TAIEX index, the Turkish BIST100 index, and the German CDAX index. In addition to the study of separate stock indices, there is also a large representation of researchers dealing with a group of stock indices. Separate stocks such as Boeing, Apple, Microsoft, TripAdvisor or a group of companies. Most researchers in this field used only one stock, or in many of them the number of test samples was insufficient, which seems to be insufficient to conclude on the impact of sentimental analysis on stock markets. Therefore far, there is no long-term research that shows remarkable results for a set of several stocks.

The vast majority of studies have focused on the US stock market. It was analyzed in a total of 55% of cases and thus became a clearly dominant market, in which the influence of sentiment analysis extracted from textual data was documented. In 15% of the work, attention was focused on the Chinese stock market and in 7% on the Taiwanese market. In terms of the European market, the German stock market appears in the outputs (7%).

Furthermore, the selected time frame selected for prediction in relevant sources is monitored and compared. Based on the analyzed sources, it is clear that the observed time period ranges from seconds to days, weeks or months. Most of the studies examined deal with daily data on stocks or indices. The researched time period ranges from several months to decades.

Table 2. Sources of numerical structured data (source: Janková, 2021)

Author	Stock	Name	Country	Period	
Hao et al. (2021)	Stock	30 stocks	Taiwan	2017	2018
Owen and Oktariani (2020)	Stock	Boeing	USA	–	2019
Sakhare et al. (2020)	Stock	n/a	n/a	n/a	n/a
Li et al. (2020)	Index	CSI 300 index	China	2009	2014
Sun et al. (2020)	Stock	n/a	China	–	2015
Nti et al. (2020)	Stock	GCB, MTNGH, TOTAL	Ghana	2010	2019
Bouktif et al. (2020)	Stock	Amazon, Apple, Microsoft, IBM, …	USA	2008	2018
Bouktif et al. (2019)	Index	CSI 300 index	USA	2007	2017
Birbeck and Cliff (2019)	Stock	Apple, Tesla, Twitter, Facebook, Netflix	USA	2015	2016
Moro et al. (2019)	Index	CSI 300 index	USA	2008	2018
Jammalamadaka et al. (2019)	Index	CSI 300 index	USA	2016	2018
Gross-Klussmann et al. (2019)	Index	ASX200, HangSeng, EURO STOXX 50	USA, EU	2010	2018
Derakhshan and Beigy (2019)	Stock	n/a	Iran, USA	2012	2013
Chen and Chen (2019)	Index	TAIEX	Taiwan	2016	2017
Ren et al. (2019)	Index	SSE 50 Index	China	2014	2016
Hwang and Kim (2019)	Stock	n/a	S. Korea	–	2019
Chen and Shih (2019)	Stock	n/a	Taiwan	–	2018

End of Table 2

Author	Stock	Name	Country	Period	
Batra a Daudpota (2018)	Stock	Apple	USA	2010	2017
Eliacik and Erdogan (2018)	Index	BIST100	Turkey	2015	2016
Kim et al. (2018)	Stock	Citigroup, Goldman Sachs, JP Morgan, ...	USA	2002	2012
Feuerriegel and Gordon (2018)	Index	DAX, CDAX, STOXX Europe 600	Germany	1996	2016
Long et al. (2018)	Index	CSI 300	China	n/a	n/a
Hajek (2018)	Stock	n/a	USA	n/a	n/a
Shi et al. (2018)	Index	CSI 300	China	2011	2015
Pagolu et al. (2017)	Stock	Microsoft	USA	2015	2016
Ab. Rahman et al. (2017)	Stock	Axiata Group, CIMB Group Holdings	Malaysia	2014	2017
Khedr et al. (2017)	Stock	Yahoo, Microsoft, Facebook	USA	n/a	n/a
Domeniconi et al. (2017)	Index	DJIA	USA	n/a	n/a
Kraus and Feuerriegel (2017)	Index	CDAX	Germany	2010	2013
Xie and Jiang (2017)	Stock	n/a	China	2008	2015
Oliveira et al. (2017)	Index	S&P 500, Nasdaq 100, Russell 2000, …	USA	2012	2015
Urolagin (2017)	Stock	n/a	n/a	–	2017
Simoes et al. (2017)	Index	S&P 500, NASDAQ 100, DJIA.	USA	n/a	n/a
Alostad and Davulcu (2017)	Index	DJIA	USA	2010	2014
Zhao et al. (2016)	Index	Shanghai Composite Index	China	–	2015
Das and Das (2016)	Stock	TripAdvisor	USA	–	2015
Eliacik and Erdogan (2016)	Index	BIST 100	Turkey	2014	2015
Hajek and Bohacova (2016)	Stock	Banks	USA	n/a	n/a
Al-Ramahi et al. (2015)	Stock	n/a	USA	2012	2014
Nguyen et al. (2015)	Stock	n/a	USA	2012	2013
Al Nasseri et al. (2015)	Index	DJIA	USA	2012	2013
Smailović et al. (2014)	Stock	Apple, Amazon, Baidu, Cisco, Google, ...	USA	–	2011
Al Nasseri et al. (2014)	Index	DJIA	USA	2012	2013
Meesad and Li (2014)	Stock	n/a	n/a	2013	2014
Tirea and Negru (2013)	Stock	n/a	Romania	2012	2013
Oliveira et al. (2013)	Stock	AMD, Amazon, Dell, Ebay, HP, …	USA	2012	2013
Nann et al. (2013)	Index	S&P 500	USA	–	2011
Siering (2012)	Index	DAX	Germany	2010	2011
O'Hare et al. (2009)	Index	S&P500	USA	–	2009

3.3. Sentiment classification algorithms

After completing the preprocessing and converting the text to several functions with numerical expression, machine learning algorithms are subsequently applied. This section provides a summary of ML algorithms that are used in the analyzed sources. The used algorithms are clearly shown in Table 3. The categorization of the reviewed studies based on the used type of algorithms is performed into 6 classes.

ML methods and ANN are widely used to classify sentiment. According to Li et al. (2020), these methods are particularly suitable for predicting highly unstable, chaotic, and volatile environments such as financial time series. If investor sentiment shows a strong predictive impulse, models with different prediction capabilities can also achieve correct results. However, if sentiment does not show a strong but weak ability to predict future developments, some models may face problems and incorrect results. If no model can be used, it probably indicates that sentiment does not have sufficient predictive power and cannot be used in practice.

The table shows that the vast majority of relevant studies and research apply the SVM method for sentiment classification. This is more than 75%, underlining the teacher learning algorithm as the mainstream of sentiment classification tools. Xie and Jiang (2017) even state that SVM shows a fantastic degree of adjustment in predicting stock fluctuations. SVM is applied in the analyzed studies either alone or in combination with other algorithms. For example, the latest study by Hao et al. (2021) uses a combination of fuzzy logic and SVM. According to the authors, fuzzy SVM is more robust compared to other methods, especially in the case of a large number of outliers. In addition, the degree of membership fuzzy decisions provide better insight into the predicted outcomes. According to the above-mentioned work, it turned out that SVM outperforms most traditional techniques in the field of stock market prediction.

The RNN type, known as Long-Short Term Memory (LSTM), is used for this purpose. In recent years, LSTMs have been very often used and promoted by networks, especially in terms of discussing the ability to predict market based on sentiment or in the case of text mining. In addition, Li et al. (2020) in their research found the superiority of the LSTM neural network in comparison with other methods and techniques in solving the problem of detecting the predictive ability of sentiment. Specifically, the authors compared LSTM with SVM methods, Naïve Bayes and logistic regression. This means that deep learning methods are a promising tool for describing the characteristics of the stock market.

On the contrary, other researchers prefer the Naïve Bayes algorithm due to its complexity. Jammalamadaka et al. (2019) incorporated online text mining into the advanced multivariate model of the Bayesian time series of machine learning, which opens the door to the simultaneous application of both text mining and machine learning in modern quantitative financial research. According to the authors of MBSTS, it surpasses models such as ARIMA, RNN or LSTM with sentimental predictors.

Table 3. Type of classification algorithm (source: Janková, 2021)

Author	SVM	ANN	Naive Bayes	Decision tree	Regression	Other
Hao et al. (2021)	x					x
Owen and Oktariani (2020)		x				x
Sakhare et al. (2020)		x		x	x	x
Li et al. (2020)	x	x	x		x	
Sun et al. (2020)	x		x			
Nti et al. (2020)		x				
Bouktif et al. (2020)	x		x			x
Bouktif et al. (2019)	x	x			x	x
Birbeck and Cliff (2019)	x		x		x	x
Moro et al. (2019)	x			x		
Jammalamadaka et al. (2019)		x				x
Gross-Klussmann et al. (2019)	x					x
Derakhshan and Beigy (2019)	x					
Chen and Chen (2019)						
Ren et al. (2019)	x					
Hwang and Kim (2019)		x			x	x
Chen and Shih (2019)	x					
Batra a Daudpota (2018)	x					
Eliacik and Erdogan (2018)	x		x			
Kim et al. (2018)	x		x		x	x
Feuerriegel and Gordon (2018)	x					
Long et al. (2018)	x					
Hajek (2018)	x			x		x
Shi et al. (2018)	x					
Pagolu et al. (2017)					x	x
Ab. Rahman et al. (2017)	x					
Khedr et al. (2017)	x		x			x
Domeniconi et al. (2017)	x			x		
Kraus and Feuerriegel (2017)	x	x		x	x	
Xie and Jiang (2017)	x					
Oliveira et al. (2017)	x	x			x	x
Urolagin (2017)	x		x			
Simoes et al. (2017)	x					x
Alostad and Davulcu (2017)	x					
Zhao et al. (2016)	x				x	
Das and Das (2016)	x		x			x
Eliacik and Erdogan (2016)	x		x			
Hajek and Bohacova (2016)	x		x	x		x

End of Table 3

Author	SVM	ANN	Naive Bayes	Decision tree	Regression	Other
Al-Ramahi et al. (2015)					x	
Nguyen et al. (2015)	x					
Al Nasseri et al. (2015)	x					
Smailović et al. (2014)	x					
Al Nasseri et al. (2014)			x			x
Meesad and Li (2014)	x					
Tirea and Negru (2013)						x
Oliveira et al. (2013)					x	
Nann et al. (2013)			x			
Siering (2012)	x					
O'Hare et al. (2009)	x		x			

The sentiment classification techniques presented in the literature have provided acceptable results in terms of the accuracy of the sentiment classification and its relationship to the stock market. A survey of text analysis concluded that the SVM and Naïve Bayes methods in particular are widely applied, while NNs are still under-explored, despite their promising potential for text classification and sentiment analysis.

4. Discussion

Based on bibliometric analysis and critical review, the findings are now synthesized and discussed in the context of leading authors and publications. It highlights the main limits and challenges in the current research identifying sentiment extracted from text messages and its impact on stock markets. The findings discussed below identify identified gaps in the research presented and may help other research teams to focus on these weaknesses and limitations and uncover new implications. Batra and Daudpota (2018) state that the initial stock market prediction research was entirely based on random walks and numerical prediction, but with the introduction of behavioral finance, people's beliefs and moods were also taken into account when predicting stock movements. Simoes et al. (2017) based on the results of the study state that it is possible to implement a profitable business strategy using textual data from social media. The empirical results of Chen and Chen (2019) suggest that big data analysis techniques to evaluate the emotional content of comments on current stocks or financial problems can effectively predict stock price movements. Al Nasseri et al. (2015) are convinced that in practice this could lead to the determination of a specific time of purchase or sale of shares and thus achieve a higher return. Such an informative decision will save investors time and effort. Shi et al. (2018) found that investor sentiment has short-term positive effects and medium-term reverse effects on the stock market. An asymmetric effect was also found, with high investor sentiment gaining a more pronounced effect on stock returns. However, Gross-Klussmann et al. (2019) still argues that there is as yet no consensus on the implementation of sentimental signals into investment strategies. Hwang and Kim (2019) found that stock prices affect sentiment, which is identified in text messages. In other words, the authors state that stock prices responded to the news before they were published. For example, Hájek (2018), who deals with a limited sample of the chosen stock market, which cannot be generalized to all markets, as recent research suggests that stock markets in other regions show specific behavior.

Another shortcoming of the above-mentioned works is the examination of only a limited period of time. Finally, there is a great shortcoming, as noted by Pagolu et al. (2016) found that the authors use data from a single source to analyze people's sentiment, but such data may be skewed because not all people who trade stocks share their views on the social platform or website in question. Oliveira et al. (2017) add that for this reason, the value of the complementarity of different data sources is not unambiguous. These sources of text data have different characteristics that can complement each other and allow more accurate predictions. For example, blogs have more complete content of opinion, microblogging content has more objectivity, interactivity and frequency of publication, and Google searches represent the opinions of more different users. A high percentage of the existing studies discussed above relied on the extraction of textual information from reputable servers or social networks Twitter or StockTweet, or from Internet discussion forums separately. Studies combining multiple text sources are available in very low proportions. However, according to the conducted secondary research, it also follows that it is appropriate to work with multiple text sources. The same opinion is held by Nti et al. (2020), who show that one of the few shows that the extraction of sentiment from combined text sets can significantly increase the accuracy of models predicting stock market developments by combining textual and numerical data. While evidence for other markets is severely limited. Based on the previous argument, it was found that the minimum researchers focused on examining the impact of sentiment in developing economies. This offers space for future research and the extension of existing knowledge to other, as yet unexplored markets.

Although it can be stated that the information contained in the extracted sentiment is effective for stock prediction on average, Nguyen et al. (2015) point out that the model with sentimental analysis shows better results for the portfolio than for individual stocks. There are many possible opinions such as that sentiment may not be a reason in stock market movements or also that sentiment can be one of the key factors that influence stock market movements, extracted sentiments from discussion forums, etc. do not reflect the price due to chaotic, erroneous commentary or prediction of human error when publishing news. In addition, all of the studies in question analyzed stocks or stock indices.

Conclusions

This paper attempted to conduct a critical review of the literature related to text mining and sentiment analysis for financial market prediction. In particular, sentiment analysis is one of the fastest growing areas of IT research, which makes it difficult to monitor all activities in this area. Although there is currently a greater amount of research that examines the use of textual data mining and the performance of sentimental analysis with consequent influence and predictive power on financial markets, only a limited number of review articles are available. The added value of this study can be seen in a systematic overview of relevant publications related to the subject matter and a detailed analysis and implications of articles, including the use of bibliometric networks and maps and the display of links between them. Not only from the bibliometric analysis, but especially from the content analysis, certain scientific gaps emerge, which are still uncovered and unresolved in this area, which provides a possible future direction and development trends of research. Relevant articles were selected from the world databases Web of Science and Scopus on the basis of a set of keywords. Out of the total number of 260 articles, a total of 49 articles meeting the specified criteria were selected for a critical overview. Bibliographic analysis was performed using VOS Viewer software. The bibliometric analysis visually mapped the past and current development of sentimental analysis in the stock markets. Through a citation analysis, the hidden links of key authors, sources and references that are key in the financial domain were revealed. And she outlined possible future developments in sentiment analysis, or in the application of expert systems in stock markets in order to predict its development. Subsequently, an extensive critical review of the relevant articles was performed in terms of the type of database, whether it is textual or numerical data. Attention was also focused on preprocessing, classification algorithms and sentiment analysis. In the end, the main findings and limits of current research were synthesized. The purpose of the presented research was to map the development of stock market sentiment analysis and to outline possible future research trends in this area through detailed content analysis, especially for researchers and scientists researching in this field, but also for laymen who are concerned about text mining and sentimental analysis as a tool for stock market prediction. The presented research has certain pitfalls and limiting shortcomings, which need to be pointed out. The selected number of analyzed papers is limited by the selected databases and the choice of keywords. Moreover, a number of articles were unavailable because they are paid platforms. Subjective is also the choice of threshold value for software debugging visualizing the bibliometric network. Proper parameter setting and map tuning requires further research in this area. In addition to the VOS Viewer software used, other tools are available. It could therefore be interesting to compare the outputs from different visualization bibliometric software and compare the results.

Academic implication

From the point of view of academic implications, the application of bibliometric analysis can be highlighted, identifying the interrelationships of selected scientific and research works in the subject, which can provide inspiration and insight into the growing popularity of stock market prediction through sentimental analysis. Including conducting a detailed critical review of international scientific and research papers, comprehensive overviews of procedures and used approaches on the basis of which the so-called white spaces are identified and thus provides an idea for the future direction of research, which can be summarized as follows:

- the outputs of individual authors often provide contradictory outputs and there is currently no consensus regarding the calculation, influence and implementation of sentiment in the context of stock markets;
- for a limited period of time;

- by one source of textual data, which distorts the extracted sentiment, or different text sources have different characteristics that can complement each other and allow for better predictions;
- the credibility and quality of text sources, as some investors may deliberately write and provide incorrect information to manipulate share prices, misinformation, sarcasm and slang terms can potentially have a very negative impact on accurate sentiment calculation;
- the vast majority of studies use only one annotated lexicon, which can bias sentiment scores;
- limited sample of stock markets that cannot be generalized to all markets as recent empirical evidence suggests that stock markets in other regions exhibit specific behavior;
- neglecting an expert model that would be able to predict the development of the stock market with the integration of sentiment with high accuracy.

Practical implication

In terms of practical implications, it can be concluded that the acquired knowledge can be used in principle by two groups of subjects. The first are investment funds, banks, central banks or research institutes, which can, by monitoring text messages, assess the mood on the stock market and monitor the connection between this mood and the movement of the market and take appropriate measures. The second group are individual investors or speculators who try to time the purchase of the stock at the most appropriate time. It would be useful for them to know to what extent it is possible to judge from the published texts in which direction and by how much the share price will change in the future, since an incorrectly determined sentiment score can subsequently distort the influence on stock markets, thus cause significant losses to investors when implemented into an investment strategy, as it may ultimately indicate incorrect buy and sell signals.

Acknowledgements

All authors who have contributed to this paper have been named, and all have no dispute about authors ordering.

Funding

This paper was supported by project No. FP-S-22-7977 “Modeling and optimization of processes in the corporate sphere” from the Internal Grant Agency at Brno University of Technology.

Author contributions

ZJ conceived the study and was responsible for the design and development of the data analysis. ZJ was responsible for data collection and analysis. ZJ was responsible for data interpretation.

Disclosure statement

Author does not have any competing financial, professional, or personal interests from other parties.

References

Ab. Rahman, A. S., Abdul-Rahman, S., & Mutalib, S. (2017). Mining textual terms for stock market prediction analysis using financial news. In *Communications in computer and information science: Vol. 788. Soft computing in data science* (pp. 293–305). Springer Verlag. https://doi.org/10.1007/978-981-10-7242-0_25

Al Nasseri, A., Tucker, A., & de Cesare, S. (2014). Big data analysis of StockTwits to predict sentiments in the stock market. In *Lecture notes in computer science: Vol. 8777. Discovery science* (pp. 13–24). Springer Verlag. https://doi.org/10.1007/978-3-319-11812-3_2

Al Nasseri, A., Tucker, A., & de Cesare, S. (2015). Quantifying StockTwits semantic terms' trading behavior in financial markets: An effective application of decision tree algorithms. *Expert Systems with Applications*, *42*(23), 9192–9210. https://doi.org/10.1016/j.eswa.2015.08.008

Al-Ramahi, M., El-Gayar, O., Liu, J., & Chang, Y. (2015). Predicting big movers based on online stock forum sentiment analysis. In *Americas Conference on Information Systems, AMCIS*.

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

Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: Current state, evolution patterns, and development priorities. *R&D Management*, *50*(3), 329–351. https://doi.org/10.1111/radm.12408

Batra, R., & Daudpota, S. M. (2018, March). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1–5). Sukkur, Pakistan. IEEE. https://doi.org/10.1109/ICOMET.2018.8346382

Birbeck, E., & Cliff, D. (2018). *Using stock prices as ground truth in sentiment analysis to generate profitable trading signals* (IDEAS Working Paper Series from RePEc). Federal Reserve Bank of St Louis. http://search.proquest.com/docview/2189119020/

Bouktif, S., Fiaz, A., & Awad, M. (2019, October). Stock market movement prediction using disparate text features with machine learning. In *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)* (pp. 1–6). Marrakech, Morocco. IEEE. https://doi.org/10.1109/ICDS47004.2019.8942303

Bouktif, S., Fiaz, A., & Awad, M. (2020). Augmented textual features-based stock market prediction. *IEEE Access*, *8*, 40269–40282. https://doi.org/10.1109/ACCESS.2020.2976725

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. *Expert Systems with Applications*, *156*, 113464. https://doi.org/10.1016/j.eswa.2020.113464

Chen, C.-H., & Shih, P. (2019, June). A stock trend prediction approach based on Chinese news and technical indicator using genetic algorithms. In *2019 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1468–1472). Wellington, New Zealand. IEEE. https://doi.org/10.1109/CEC.2019.8790177

Chen, M.-Y., & Chen, T.-H. (2019). Modeling public mood and emotion: Blogs and news sentiment and socio-economic phenomena. *Future Generation Computer Systems*, *96*, 692–699. https://doi.org/10.1016/j.future.2017.10.028

Das, S., & Das, A. (2016, July). Fusion with sentiment scores for market research. In *FUSION 2016 – 19th International Conference on Information Fusion, Proceedings* (pp. 1003–1010). Heidelberg, Germany. IEEE.

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

Domeniconi, G., Moro, G., Pagliarani, A., & Pasolini, T. (2017). Learning to predict the stock market dow jones index detecting and mining relevant tweets. In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. (pp. 165–172). Funchal, Madeira, Portugal. SciTePress. https://doi.org/10.5220/0006488201650172

Eliacik, A., & Erdogan, N. (2015, November). User-weighted sentiment analysis for financial community on Twitter. In *2015 11th International Conference on Innovations in Information Technology (IIT), Conference Proceedings* (pp. 46–51). Dubai, United Arab Emirates. IEEE. https://doi.org/10.1109/INNOVATIONS.2015.7381513

Eliacik, A. B., & Erdogan, N. (2018). Influential user weighted sentiment analysis on topic based microblogging community. *Expert Systems with Applications*, *92*, 403–418. https://doi.org/10.1016/j.eswa.2017.10.006

Feuerriegel, S., & Gordon, J. (2018). Long-term stock index forecasting based on text mining of regulatory disclosures. *Decision Support Systems*, *112*, 88–97. https://doi.org/10.1016/j.dss.2018.06.008

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

Hájek, P. (2018). Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. *Neural Computing and Applications*, *29*(7), 343–358. https://doi.org/10.1007/s00521-017-3194-2

Hájek, P., & Boháčová, J. (2016). Predicting abnormal bank stock returns using textual analysis of annual reports – A neural network approach. In *Communications in computer and information science: Vol. 629*. *Engineering applications of neural networks* (pp. 67–78). Springer, Cham. https://doi.org/10.1007/978-3-319-44188-7_5

Hao, P. Y., Kung, C. F., Chang, C. Y., & Ou, J. B. (2021). Predicting stock price trends based on financial news articles and using a novel twin support vector machine with fuzzy hyperplane. *Applied Soft Computing*, *98*, 106806. https://doi.org/10.1016/j.asoc.2020.106806

Huang, C., Liao, J., Yang, D., Chang, T., & Luo, Y. (2010). Realization of a news dissemination agent based on weighted association rules and text mining techniques. *Expert Systems with Applications*, *37*(9), 6409–6413. https://doi.org/10.1016/j.eswa.2010.02.078

Hwang, E., & Kim, Y. (2019). *Interdependency between the stock market and financial news*. ArXiv. https://doi.org/10.48550/arXiv.1909.00344

Jammalamadaka, S., Qiu, J., & Ning, N. (2019). Predicting a stock portfolio with the multivariate bayesian structural time series model: Do news or emotions matter? *International Journal of Artificial Intelligence*, *17*(2), 81–140. https://escholarship.org/uc/item/47m0302b

Janková, Z. (2021). *Expert system for decision-making on stock markets using investor sentiment* [Doctoral Thesis]. Brno University of Technology. Faculty of Business and Management.

Khedr, A., Salama, S. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, *9*(7), 22–30. https://doi.org/10.5815/ijisa.2017.07.03

Kim, M., Park, E. L., & Cho, S. (2018). Stock price prediction through sentiment analysis of corporate disclosures using distributed representation. *Intelligent Data Analysis*, *22*(6), 1395–1413. https://doi.org/10.3233/IDA-173670

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems, 104*(C), 38–48. https://doi.org/10.1016/j.dss.2017.10.001

Li, Y., Bu, H., Li, J., & Wu, J. (2020). The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. *International Journal of Forecasting*, *36*(4), 1541–1562. https://doi.org/10.1016/j.ijforecast.2020.05.001

Long, W., Tang, Y. R., & Tian, Y. J. (2018). Investor sentiment identification based on the universum SVM. *Neural Computing and Applications*, *30*(2), 661–670. https://doi.org/10.1007/s00521-016-2684-y

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

Meesad, P., & Li, J. (2014, December). Stock trend prediction relying on text mining and sentiment analysis with tweets. In *2014 4th World Congress on Information and Communication Technologies (WICT 2014)* (pp. 257–262). Melaka, Malaysia. https://doi.org/10.1109/WICT.2014.7077275

Moro, G., Pasolini, R., Domeniconi, G., Pagliarani, A., & Roli, A. (2019). Prediction and trading of dow jones from Twitter: A boosting text mining method with relevant tweets identification. In *communications in computer and information science: Vol. 976. Knowledge discovery, knowledge engineering and knowledge management* (pp. 26–42). Springer, Cham. https://doi.org/10.1007/978-3-030-15640-4_2

Nann, S., Krauss, J., & Schoder, D. (2013). Predictive analytics on public data – The case of stock markets. *ECIS 2013 Completed Research.* 102.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603–9611. https://doi.org/10.1016/j.eswa.2015.07.052

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). Predicting stock market price movement using sentiment analysis: Evidence from Ghana. *Applied Computer Systems*, *25*(1), 33–42. https://doi.org/10.2478/acss-2020-0004

O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., &. Smeaton, A. F. (2009, November). Topic-dependent sentiment analysis of financial blogs. In *Proceeding of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion – TSA '09* (pp. 9–16). Hong Kong China. Association for Computing Machinery. https://doi.org/10.1145/1651461.1651464

Oliveira, N., Cortez, P., & Areal, N. (2013, June). Some experiments on modeling stock market behavior using investor sentiment analysis and posting volume from Twitter. In *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics – WIMS '13* (pp. 1–8). https://doi.org/10.1145/2479787.2479811

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, *73*(C), 125–144. https://doi.org/10.1016/j.eswa.2016.12.036

Owen, L., & Oktariani, F. (2020, August). SENN: Stock ensemble-based neural network for stock market prediction using historical stock data and sentiment analysis. In *2020 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 1–7). Bandung, Indonesia. IEEE. https://doi.org/10.1109/ICoDSA50139.2020.9212982

Pagolu, V., Kamal, N. R. C., Panda, G., & Majhi, B. (2016). *Sentiment analysis of Twitter data for predicting stock market movements.* ArXiv. https://doi.org/10.48550/arXiv.1610.09225

Ren, R., Wu, D. D., & Liu, T. (2019). Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*, *13*(1), 760–770. https://doi.org/10.1109/JSYST.2018.2794462

Sakhare, N. N., Imambi, S., Kagad, S., Kapadwanjwala, T., Malekar, M., & Dalal, M. (2020). Stock market prediction using sentiment analysis. *International Journal of Advanced Science and Technology*, *29*(4s), 1126–1133. http://sersc.org/journals/index.php/IJAST/article/view/6664

Shi, Y., Tang, Y.-R., Cui, L.-X., & Long, W. (2018). A text mining based study of investor sentiment and its influence on stock returns. *Economic Computation and Economic Cybernetics Studies and Research*, *52*(1), 183–199. https://doi.org/10.24818/18423264/52.1.18.11

Siering, M. (2012, January). "Boom" or "Ruin" – Does it make a difference? Using text mining and sentiment analysis to support intraday investment decisions. In *2012 45th Hawaii International Conference on System Sciences* (pp. 1050–1059). Maui. IEEE. https://doi.org/10.1109/HICSS.2012.2

Simoes, C., Neves, R., & Horta, N. (2017, June). Using sentiment from Twitter optimized by Genetic Algorithms to predict the stock market. In *2017 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1303–1310). Donostia, Spain. IEEE. https://doi.org/10.1109/CEC.2017.7969455

Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2014). Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences*, *285*(C), 181–203. https://doi.org/10.1016/j.ins.2014.04.034

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. https://doi.org/10.1002/asi.4630240406

Sun, Y., Liu, X., Chen, G., Hao, Y., & Zhang, Z. (2020). How mood affects the stock market: Empirical evidence from microblogs. *Information* & *Management*, *57*(5), 103181. https://doi.org/10.1016/j.im.2019.103181

Tirea, M., & Negru, V. (2013, September). Classifying and quantifying certain phenomena effect. In *SISY 2013 – IEEE 11th International Symposium on Intelligent Systems and Informatics, Proceedings* (pp. 363–368). Subotica, Serbia. IEEE. https://doi.org/10.1109/SISY.2013.6662603

Urolagin, S. (2017). Text mining of Tweet for sentiment classification and association with stock prices. In *2017 International Conference on Computer and Applications (ICCA)* (pp. 384–388). Doha, Qatar. IEEE. https://doi.org/10.1109/COMAPP.2017.8079788

van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Xie, Y., & Jiang, H. (2017). Stock market forecasting based on text mining technology: A support vector machine method. *Journal of Computers*, *12*(6), 500–510. https://doi.org/10.17706/jcp.12.6.500-510

Zhao, B., He, Y., Yuan, C., & Huang, Y. (2016, July). Stock market prediction exploiting microblog sentiment analysis. In *2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 4482–4488). Vancouver, BC, Canada. IEEE. https://doi.org/10.1109/IJCNN.2016.7727786