Review Article

Sentiment Analysis Using Machine Learning in Stock Market: A Bibliometric Visualization

Vision
1–18

DOI: 10.1177/09722629231172580
journals.sagepub.com/home/vis
Sage

Shivani[1] and Bhag Singh Bodla[1]

Abstract

Sentiments, opinions and emotions are considered as the basis for various kinds of analysis and anticipations. Sentiment analysis as a concept of generating the polarity from written, visual and verbal content is burgeoning in the advanced technological era. The article aims to carry out a bibliometric visualization study that reviews the current status, worldwide collaboration, evolution and potential emergence of machine learning-based sentiment analysis in the area of financial market along with several other themes. A corpus of 610 research publications from the Web of Science database during 2008–2022 is the base for depicting the contribution of eminent authors, journals, countries and research themes. Outcomes of the study reveal the wide scope of the 'Sentiment Analysis' theme in the existing research world along with providing a direction for future studies which may embed several advanced topics. The theme-related publications are segmented into four clusters where the 'sentiment analysis and predictions using machine learning' reflects the future emergence with existing growth. This concept has widely spread in India, China and the USA. Potential researchers can implement this concept in the stock market and financial predictions, sarcasm detection and behavioural mapping as major contributions to the firms' effective decisions.

Key Words

Sentiment Analysis, Stock Market Prediction, Machine Learning, Natural Language Processing, Twitter

Introduction

It needs mention that in the epoch of advanced computing and information technology, the behavioural aspects of human beings occupy a significant place. Along with several behavioural biases, the concept of sentiments and opinions has revolutionized in the twenty-first century as the key tool to analyse the moods and beliefs of people. Liu (2012) defines 'sentiment' as a quintuple (combination of five words, that is E, A, S, H, T), where 'E' depicts an entity, 'A' presents an aspect, 'S' reflects sentiment on 'E' entity's 'A' aspect, 'H' is the holder of opinion and 'T' refers to the time of opinion expression. Pang et al. (2002) propounded the concept of sentiment analysis to classify movie reviews. Sentiment analysis is the technique to get polarity from people's emotions, moods and opinions through texts and other sets of information (Tang et al., 2015). In sentiment classification and polarity generation, the sentiments can be positive, negative or neutral which gives a refined set of views (Feldman, 2013).

In early twenty-first century, this concept was initially applied to classify movie reviews (Casoto et al., 2008; Pang et al., 2002), news documents (Shi et al., 1992) and other text content using library-based techniques (Sul et al., 2017), manual methods and traditional unsupervised lexicon models (Sakhare et al., 2020). Earlier the machine learning-driven sentiment analysis was performed using traditional models (Agarwal et al., 2021; Al-Rubaiee et al., 2016; Smailović et al., 2013) but now the sentiment analysis can be done through artificial neural networks, natural language processing technique, maximum entropy, supervised lexicons and other python based tools (Kinyua et al., 2021; Mcgurk et al., 2020; Mendoza-Urdiales et al., 2022; Valle-Cruz et al., 2022; Xu et al., 2022).

In the present era, the concept of sentiment analysis has become indispensable in the field of finance and other areas (Grljević et al., 2022; Krishnamoorthy, 2018; Vermeer et al., 2019). In finance, the concept is largely used to analyse the impact of human sentiments' on stock market returns (Tabari et al., 2019; Yang et al., 2015) along

[1] University School of Management, Kurukshetra University, Kurukshetra, Haryana, India

Corresponding author:
Shivani, University School of Management, Kurukshetra University, Kurukshetra, Haryana 136119, India.
E-mail: ssaini31998@gmail.com

with various other financial domains. Various studies show that public sentiments affect stock prices and returns in the future (Bollen et al., 2011; Carosia et al., 2020; Reboredo & Ugolini, 2018; Yılmaz, 2022). Analysis of sentiments and psychological behaviour aids investors and market participants to make their prospective strategies (Valle-Cruz et al., 2022).

Few bibliometric studies have been conducted which show the evaluation of sentiment analysis. Ahlgren (2016) and Keramatfar and Amirkhani (2019) investigated the evolution of the sentiment analysis concept. Li and Lei (2021) evaluated the development of a specific text mining tool 'Topic Modeling' from 2000 to 2017. Cobo (2020) measured the trend of sentiment analysis and opinion mining. Lopez-Martinez and Sierra (2021) studied the trend of sentiment analysis technique 'Natural Language Processing' in Mexico.

After scanning the existing literature, we have found that a few bibliometric studies deal with the sentiment analysis concept, but rarely any of them is related to machine learning-driven sentiment analysis in stock market field. Our study aims to cover this research gap by answering the following questions:

1. What is the existing trend of machine learning-based sentiment analysis in stock market?
2. What is the status of thematic evolution and worldwide research collaboration for this theme?
3. What are the emerging topics that can be embedded with this theme?

The purpose to conduct this research is to provide a trend evaluation of sentiment analysis concept in the field of stock market and other financial areas. It contributes to the literature by executing bibliometric descriptive and network analysis, conceptual evaluation and thematic evolution of the concept of machine learning-based sentiment analysis in stock market. It presents a deep scanning of the publications in the domain of sentiment analysis using machine learning from 2008 to mid-2022.

This research is expected to benefit future researchers to identify the most prolific authors, countries and publication sources which are providing a light to the underlying theme of the present study. Additionally, it will aid the corporates through analysing issues which can be solved through machine learning-driven sentiment analysis.

In remaining article, the second section covers the existing review of literature. The third section covers the methodology used to extract the corpus and perform the bibliometric analysis. Furthermore, the fourth section presents results and interpretations of the analysis. The fifth section illustrates the findings of the study, concept framework, contribution, limitations and future directions. Eventually, the sixth section presents the conclusion of the research study.

Related Literature

Initially, studies in the field of sentiment analysis were performed for written content and reviews. Casoto et al. (2008) and Pang et al. (2002) explored this concept for reviewing movie watchers' reviews through machine learning models. Further, Edmans et al. (2007) conducted a study related to emotional sentiment detection for sports results. In the field of information technology, Uhl and Ag (2014) took Reuters' news articles to classify the sentiments of news published in articles. Extending it, the detection of news-driven sentiments was performed by Vijay et al. (2018). Yadav et al. (2020) also considered financial news text to analyse the sentiments and segregated views of published news through unsupervised lexicon techniques. Further, Rao and Srivastava (2012), Teti et al. (2019) and Mcgurk et al. (2020) generated the sentiment polarity of investors' posted tweets through various machine learning techniques. Grljević et al. (2022) applied the machine learning-based classification concept to annotate the opinions in the area of higher education through social media websites.

With the advent of advanced technology, various related research have been performed by researchers through supervised models. Mohamed (2017) and Catal and Nangir (2017) observed that support vector machine (SVM) sentiment classifiers outperform the unsupervised models such as traditional naive Bayes, K-nearest neighbours and the decision-tree model. Tan and Tas (2021) measured the investors' sentiments through the SVM. Singh and Singh (2021) investigated the effectiveness of 10 off-the-shelf techniques and 17 machine learning tools used for public opinion mining against various societal issues. Furthermore, Kinyua et al. (2021) applied a python-based valence aware dictionary for sentiment reasoning technique to classify the sentiments of tweets posted by the US President Donald Trump during the US presidential elections. Valle-Cruz et al. (2022) implemented sentic computing, supervised lexicons and natural language processing technique to detect the sentiments of influential accounts during global pandemics.

The concept of sentiment analysis has widely applied in the field of financial market forecasting. Smailović et al. (2013) showed the predictive impact of sentiment polarity (on stock price movements) and found that the sentiments can forecast the price movements a few days in advance. Kinyua et al. (2021), Hassanein et al. (2021) and Oliveira et al. (2017) found that researchers can predict stock prices through the sentiments of investors and other influential personalities. Furthermore, the association between tweet sentiments and trading volume along with stock returns has been found by Tan and Tas (2021). Sakhare et al. (2020), Leitch and Sherif (2017), Zhang et al. (2011) and Garcia-Lopez et al. (2018) found the significantly negative association between tweet sentiments and stock price movements in future and Xiao Li et al. (2017) presented a positive link between daily happiness sentiment and stock returns in the future days. Saurabh and Dey (2020) observed an impact of social dimensions on predictive stock returns. Further, Pagolu et al. (2016) have shown that public sentiments can strongly affect the rise and fall in the stock prices.

Bibliometric Materials and Methods

The present research employed a specific methodology in different phases, that is planning and data capturing from a database, analysis and reporting of research outcomes.

Data Capturing from Bibliographic Databases

At the stage of data capturing, the study chose to extract data from the Web of Science database for retrieving the relevant research articles for 15 years period. A document-based search was executed on 23rd June 2022 in the Web of Science database for the exploration of research articles using the search query 'Sentiment Analysis' AND ('Machine Learning' OR 'Stock Market Prediction'). This keyword enabled the extraction of 1,193 publications since 2008 without using any filter. Further, the search query 'Sentiment Analysis' AND ('Machine Learning' OR 'Stock Market Prediction') (Topic) and English (Language) and Computer Science Information System or Computer Science Artificial Intelligence or Business or Management or Business Finance (Web of Science Categories) was filled which revealed 712 records. After excluding paper proceedings, early access papers, letters and other non-full text articles, the final corpus of 610 research articles was extracted. To explore the application of sentiment analysis in the field of finance, business and advanced computing, the collection of 610 full-text research articles was used for bibliometric descriptive analysis and network visualization from 2008 to 2022.

Framework for Anatomizing the Data

The analysis of this article was performed through the most relevant components, that is author analysis, source analysis, country analysis and theme analysis. At first, description analysis was performed from 2008 to 2022 on the corpus of 610 research documents by taking the top 10 authors, sources, countries and themes using MS Excel 2019 and 'R' software package 'Biblioshiny' given by Aria and Cuccurullo (2017). Network clustering and visualization were executed using the bibliometric software 'VOSviewer' 1.6.18 which was propounded by van Eck and Waltman (2017).

Figure 1 presents the research framework which contains extraction and analysis phases for bibliometric visualization.

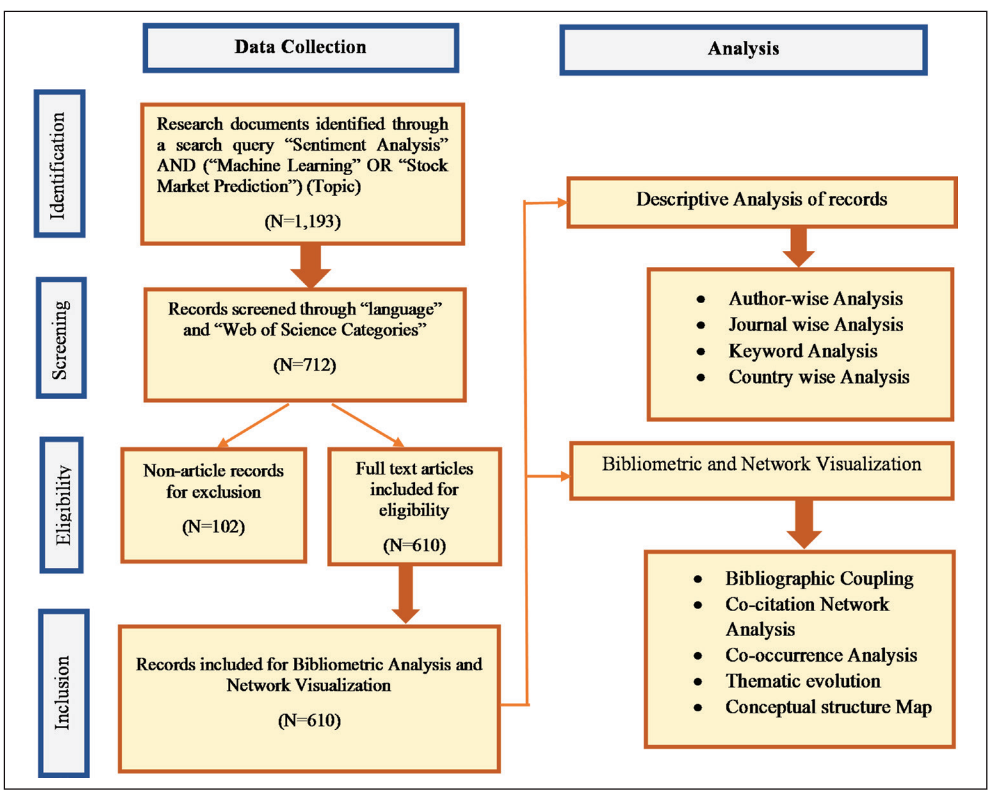

Figure 1. Research Framework.

Results and Discussion

Results

This section elucidates the descriptive analysis of research documents related to sentiment analysis implementation in the context of stock market and other prediction.

Descriptive Analysis of Authors, Countries, Journals and Keywords

A descriptive analysis presents several quantitative information regarding documents trend, citations, authors, collaboration and keywords. Table 1 depicts the analysis of documents.

Among 610 research documents, 21 documents are single-authored and the remaining 589 are multi-authored documents which published in 150 journals. These are published by 1,938 authors where 1,920 authors published the articles with mutual collaboration.

Table 1. Descriptive Analysis of Documents Related to Research Theme.

Description	Results
Total documents	610
Period	2008–2022
Sources (journals, books and others)	150
Average citations per document	2.79
Authors	1,938
Author appearances	2,268
Authors of single-authored documents	18
Authors of multi-authored documents	1,920
Single-authored documents	21
Documents per author	0.315
Authors per document	3.18
Co-authors per documents	3.72
Collaboration index	3.26
Keywords plus (ID)	578
Author's keywords (DE)	1,589

A collaboration index of 3.26 shows that more than three authors have published a paper in the case of one multi-author document. Documents per author or the ratio between total documents and total authors (610/1,938) shows that only 0.31347 portion of a document is published by an author. Furthermore, the authors per document or the ratio between total authors and total documents (1,938/610) is 3.17704 which presents that one paper is published by more than three authors. The average citations per document referring the number of times an external paper cites a document. The citations are 2.79 per document which implies that one document has referred by 2.79 local or global documents.

To get a synthesis of research documents published, a trend analysis is plotted vis-à-vis document publications from 2008 to 2022 (June) in Figure 2.

Figure exhibits an upward publication trend of the documents from 2012 to 2021. During 2008 and 2011, this theme was not much popular due to infancy stage of AI, big data and deep learning. Gradually, the growth of machine learning and sentiment analysis concept persuaded the researchers to publish documents in this field. In 2021, 153 research documents were published which show a growth rate of 206% in comparison to 2018. In 2022, up to June, 79 papers have been published by authors from different countries in this expertise. Hike in the published documents from the past to present indicates that the concept of sentiment analysis is widely analysed using different ML models in the domain of financial markets.

Country-wise Annotation

A country-wise examination through Table 2 shows that the three most valuable countries in terms of total citations (TC) are India with 1,812 citations, China with 1,678 and the United Kingdom with 1,397 citations followed by Spain and the USA. India occupies 19.95% portion of the top 10 countries' TCs. The TCs illustrate the citations for

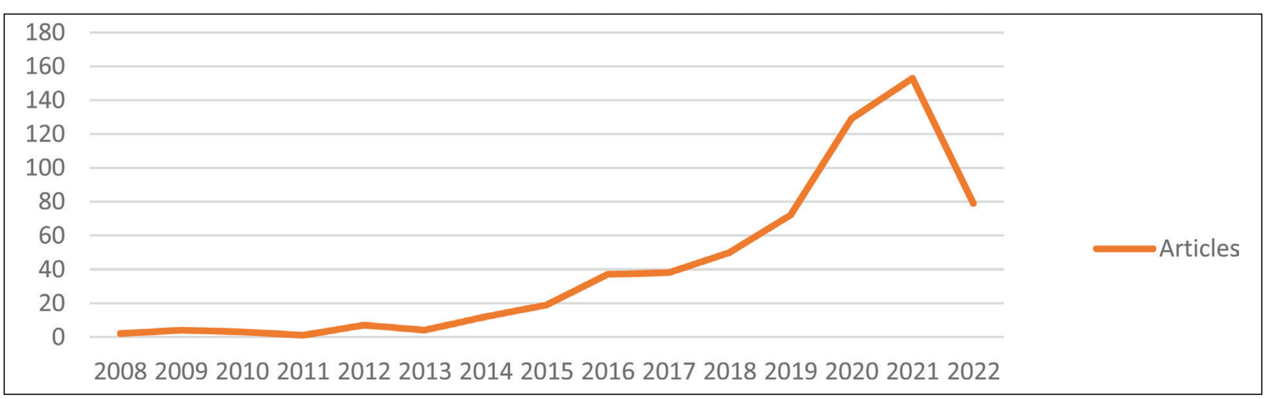

Figure 2. Publication Trend of Research Documents.

Source: Web of Science, Search Date: 23rd June 2022.

overall research documents published in a country over an entire study period. Turkey and Belgium are the least cited nations whose TCs occupy only 4.81% and 4.44% of the top 10 countries' overall citations. Additionally, in terms of per publication citations (C/P), Belgium is the most cited country.

The most productive countries in terms of the total volume of publications (TP) in this field are China, India and the USA followed by Pakistan and the United Kingdom. Turkey and Italy are the least contributing countries with 44 total publications between 2008 and 2022 (Figure 3).

To get an understanding of the most effective country, C/P is the most reliable measure (Zuopeng et al., 2021). In this term, Belgium is the most effective nation around the world because the publications from this country have referred by maximum authors, whereas China is the highest contributing country because the author from this nation has researched on a wide scale related to different applications of sentiment analysis concept.

Table 2. Top 10 Countries Based on Per Paper and Total Citations.

Country	Total Citations (TC)	% of TC	Citations Per Paper
India	1,812	19.95	11.92105263
China	1,678	18.48	7.524663677
United Kingdom	1,397	15.38	19.40277778
Spain	858	9.45	14.54237288
USA	847	9.33	7.7
Singapore	632	6.96	27.47826087
Greece	532	5.86	23.13043478
Canada	486	5.35	16.75862069
Turkey	437	4.81	9.931818182
Belgium	403	4.44	44.77777778
Total	9,082	100.00	183.1677793

The descriptive analysis depicts that the 589 research documents are multi-authored which include the collaboration of multiple countries' authors. Based on the total publication of author countries, top collaborative countries are Pakistan and Saudi Arabia with 15 joint publications. Authors from Pakistan also published nine joint papers with Korean authors in the field of sentiment analysis with machine learning models (see Figure 4).

Authors from China present a high-to-moderate degree of collaboration with the USA, Singapore and the United Kingdom with 14, 8 and 7 publications. The USA shows more synergy with China, whereas less mutual authorship with Korea. Furthermore, India does not show any cooperation with the USA in terms of combined research publication as per the top 10 collaborative countries.

In nutshell, the high publication in the field of sentiment analysis and machine learning in China, India, and the USA is the result of advancement of big data and information technology.

Journal or Source-wise Annotation

The analysis of the top 10 journals based on TCs and total production is illustrated in Table 3. Furthermore, Figure 5 explains the year-wise growth of the top 10 journals during the study period.

Among a corpus of journals, *Expert Systems with Applications*, *Knowledge-Based Systems*, *IEEE Access*, *Information Processing & Management*, *Journal of Information Science*, *Cognitive Computation*, *Soft Computing*, *Multimedia Tools and Applications*, *Electronics* and *CMC—Computers Materials & Continua* are the top sources citing the documents in different field of sentiment analysis associated with machine learning techniques and stock market forecasting. Among these top journals, the first rank journal plays a dominant role in terms of citations

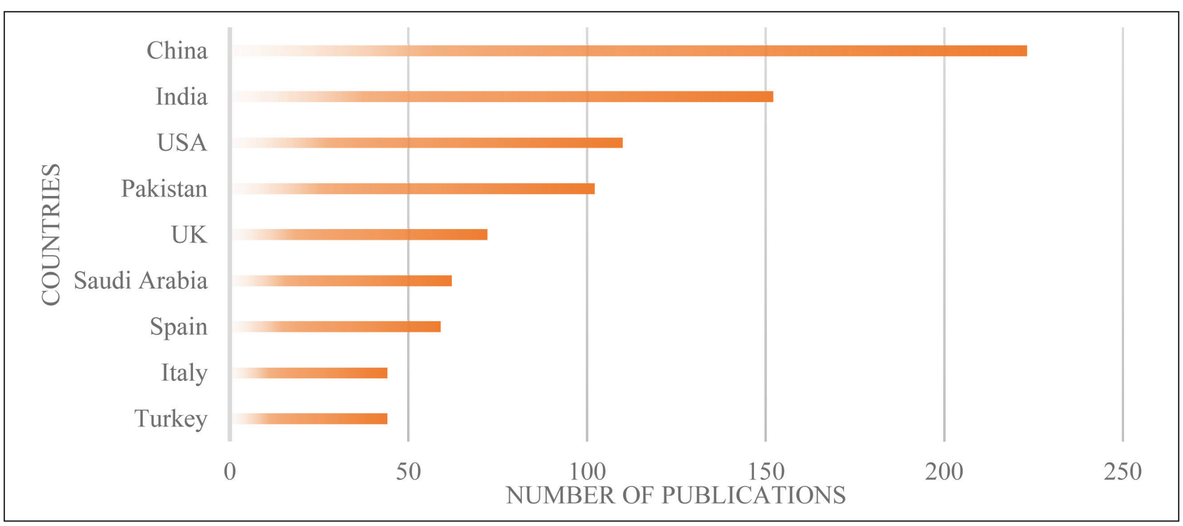

Figure 3. Top 10 Countries Based on Total Publication Production.

Source: Web of Science, Search Date: 23rd June 2022.

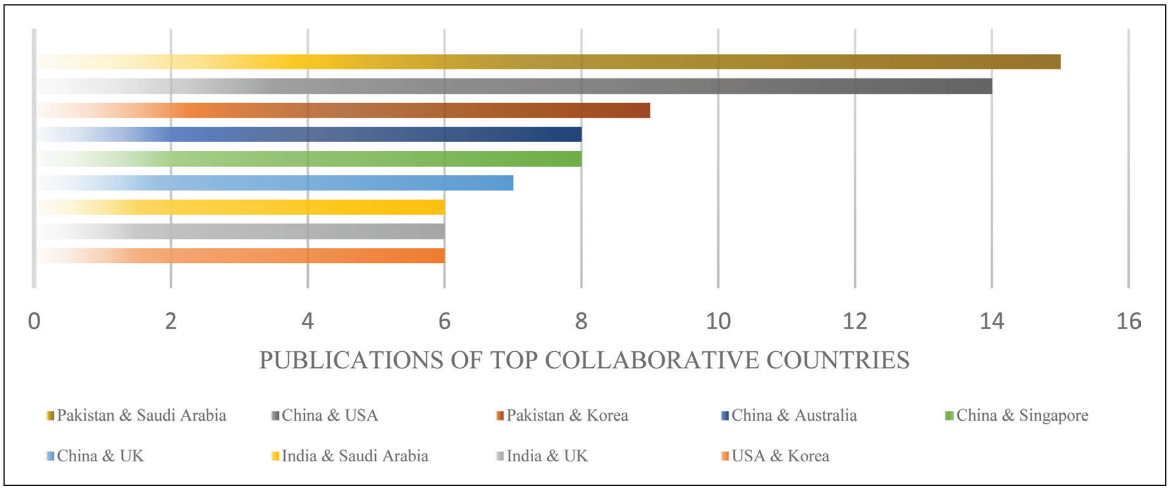

Figure 4. Top 10 Countries Based on Research Paper Collaboration.

Source: Web of Science, Search Date: 23rd June 2022.

Table 3. Top 10 Journals Based on Total Production and Total Citation.

Sources	TP	TC
Expert Systems with Applications	42	2,104
Knowledge-Based Systems	31	1,561
IEEE Access	91	936
Information Processing & Management	22	741
Journal of Information Science	14	385
Cognitive Computation	17	293
Soft Computing	13	137
Multimedia Tools and Applications	14	118
Electronics	15	106
CMC—Computers Materials & Continua	19	98

(2,104) and publications (42). Except *CMC—Computers Materials & Continua*, all the above journals have been cited by more than 100 authors.

In terms of publication of articles since inception, *IEEE Access* is the prolific source with 91 total publications followed by *Expert Systems with Applications* and *Knowledge-Based Systems* with 42 and 31 publications, respectively. As *IEEE Access* started publishing research papers in 2016, its publications show a sharp growth up to June 2022 as compared to other journals (Figure 5). *IEEE Access* has a sharp rise in terms of publications from 2018 to 2022 because the authors have published a large number of papers related to sentiment analysis concept in stock market and machine learning applications in stock market. Most of the journals started publications in 2014, whereas the top-cited journal shows the publications from 2008. *Soft Computing* is the least productive source as it has published only 13 research papers up to 2022. These journals promote the research related to machine learning and advance computing concepts which enabled high publication of documents in these themes.

Author-wise Annotation

A descriptive analysis was carried (Table 4) to get insights about the collaborations between researchers, relevance in terms of their cited documents and publications in various domains.

Among top productive authors, Ravi Kumar (Ravi in Table 4) from India is found as the most cited author with the highest C/P of 561 followed by Kevan Buckley (Buckley, K.) from England, Ravi Vadlamani (Ravi, V.) from India and Georgious Paltoglou (Paltoglou, G.) from England with a C/P of 537, 334.5 and 320, respectively. Except Erik Cambria, majority of the top authors have published less than five papers. In terms of TCs and publications, Erik Cambria (Cambria, E.) is found as the top author with the largest publication (12 publications) among all the top effective authors. He has an expertise in the domain of artificial intelligence and machine learning models.

Authors with high average citations (C/P) do not present high publications indicating that the theme of sentiment analysis and stock market prediction is in trend. Upcoming authors are highly referring to the publications related to these domains for conducting the future research.

Theme or Keyword-based Analysis

To be conversant with the emerging and dominant research themes in the area of sentiment analysis, a keyword-based anatomization is effectuated. Figure 6 illustrates a word cloud based on the top 30 keywords of the research corpus. Among all themes, text and option classification along with sentiment analysis are the most frequent topics in the field of finance, computer science and other management expertise (Ajitha et al., 2021; Lin et al., 2020).

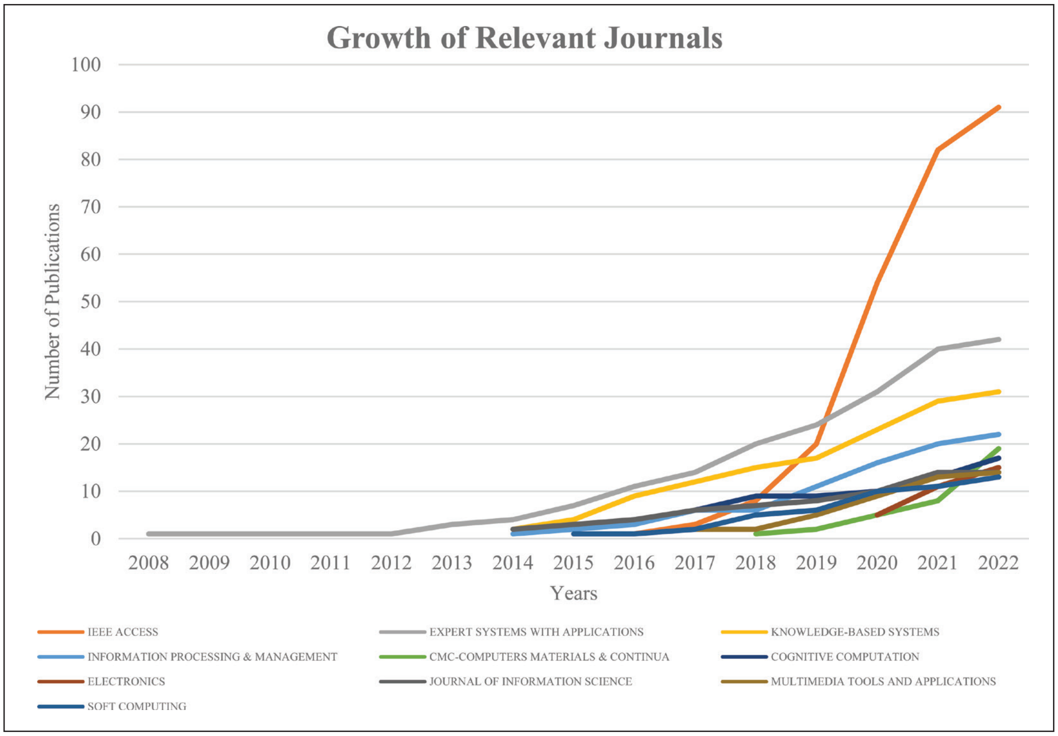

Figure 5. Growth Dynamics of Top 10 Productive Journals.

Source: Web of Science, Search Date: 23rd June 2022.

Table 4. Top 10 Authors on the Basis of Total Citations and Overall Contribution.

Authors	Countries	TC	C/P	TP	*m* Index	*h* Index	*g* Index
Cambria, E.	Singapore	788	65.66667	12	10	12	1.111
Ravi, V.	India	669	334.5	2	2	2	0.25
Thelwall, M.	England	669	167.25	4	4	4	0.364
Paltoglou, G.	England	640	320	2	2	2	0.182
Ravi, K.	India	563	563	1	1	1	0.125
Buckley, K.	England	537	537	1	1	1	0.091
Poria, S.	Mexico	339	113	3	3	3	0.333
Onan, A.	Turkey	331	110.3333	3	3	3	0.429
Wu, Dd.	USA	321	160.5	2	2	2	0.154
Korukoglu, S.	Turkey	311	155.5	2	2	2	0.286

Notes: TC: Total citations, C/P: citation per paper, TP: total publication.

Model and impact keywords express the use of unsupervised and supervised machine learning models to measure the impact of sentiments on stock market movements and other business decisions (Bouktif et al., 2020; Picasso et al., 2019). Additionally, social media, Twitter, reviews, sentiment, feature selection and networks are the major keywords which apply in advanced financial and computer science research.

Bibliographic Coupling of Journals, Authors and Countries

To examine the mutual linkage and collaboration, the bibliometric coupling was conducted which shows the common citation used by two more documents and authors (Figure 7). As per the bibliographic coupling of journals having more than 100 citations, the journal *IEEE Access* has maximum links with other journals. It depicts that the papers published on a theme of sentiment analysis and stock market prediction in other journals have a link with any of the 91 documents listed in *IEEE Access* (see Table 5). In terms of period analysis, *IEEE Access*, *Multimedia Tools and Applications*, *Journal of Business Research*, *Knowledge-Based Systems*, *Expert System with Applications*, *International Journal of Research* in *Marketing and Electronics* have high mutual links from 2018 to 2022 which presents the interdependence of these journals in this domain.

Figure 6. Theme Analysis Based on Word Cloud Framework.

Source: Biblioshiny.

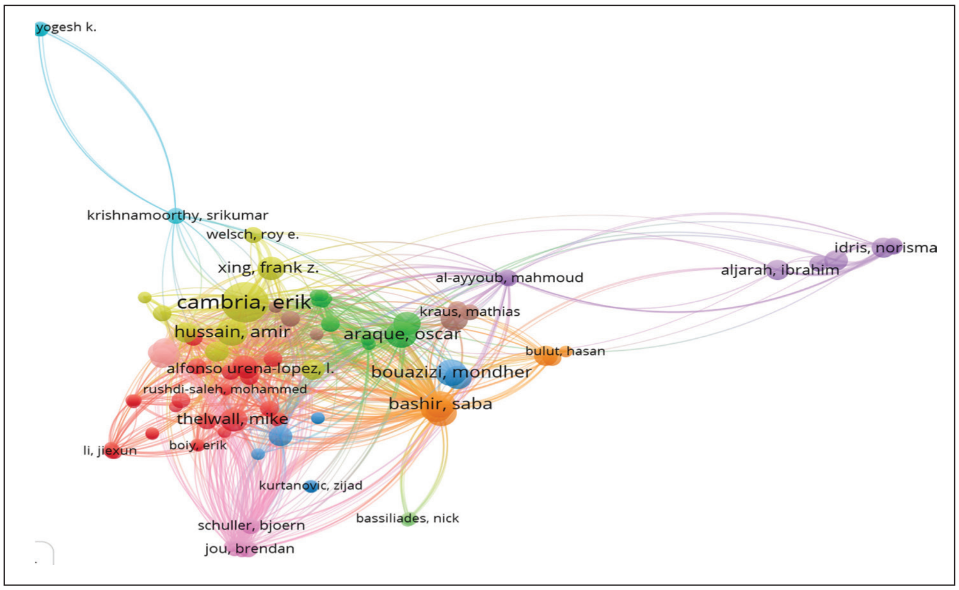

Figure 7. Bibliographic Coupling Based on Authors.

Source: Vosviewer.

Table 5. Top 10 Authors in Form of Bibliometric Coupling.

Authors	Total Link Strength
Cambria, Erik	3,563
Khan, Farhan Hassan	3,019
Qamar, Usman	3,019
Ravi, V	1,896
Alfonso Urena-Lopez, L.	1,885
Teresa Martin-Valdivia, M.	1,668
Araque, Oscar	1,668
Schuller, Bjoern	1,607

Figure 8 shows the coupling of top-linked authors which shows that Erik Cambria, Oscar Araque, Amir Hussain, Mike Thelwall, Frank Z. Xing and Saba Bashir are highly linked due to mutual citations. Erik Cambria, Farhan Hassan Khan and Usman Qamar are the top authors with high bibliometric links. Norisma Idris, Mahmoud Al-Ayyouub and Oscar Araque are from the same cluster and actively collaborative authors.

In terms of country-wise bibliographic coupling, India has the highest link strength with other countries followed by China and the USA (Table 6). It expresses that the documents published in other countries are highly cited the publications from these countries. India, Canada, the USA and Germany present high cooperation in terms of publications on a theme of sentiment analysis through machine learning along with stock market prediction through sentiment analysis (Figure 8).

Co-authorship Analysis

As per the descriptive analysis, 1,920 authors have published the paper in collaboration with other authors. Among all, Erik Cambria has the largest network in terms of co-authorship with 10 other authors from different countries.

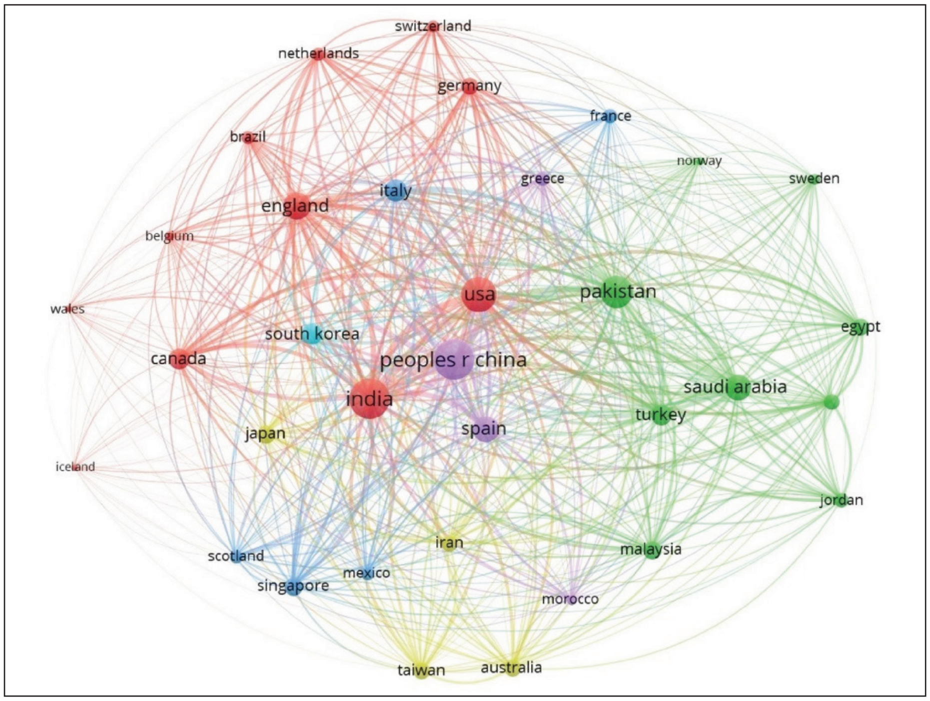

Figure 8. Bibliographic Coupling Based on Countries.

Source: Vosviewer.

Table 6. Top 10 Countries on the Basis of Bibliometric Coupling.

Countries	Total Link Strength
India	34,983
USA	32,059
China	31,237
Pakistan	20,112
England	19,429
Spain	17,093
Saudi Arabia	11,414
Italy	10,964
Canada	10,716
Iran	9,208

Figure 9 shows that Cambria published 12 research documents with Feng Xu, Frank Z. Xing, Roy E. Weisch, Gregoire Winterstein, Soujanya Poria, Gaung Bin Huang, Alexander Gelbukh, Amir Hussain, Huan Zhao and Yunqing Xia on various themes of sentiment analysis and stock market prediction using several advanced algorithmic models.

Among all authors, these are largely connected authors and published the research related to sentiment polarity determination through advanced techniques along with its application in different fields, that is financial market movement prediction, asset allocation and psychological common-sense reasoning (Lin et al., 2020; Oneto et al., 2017; Picasso et al., 2019; Poria et al., 2014; Xia et al., 2016; Xing et al., 2018).

Co-citation Analysis

The analysis of co-citations for research publications assists in providing the knowledge of mutual interdependence among authors, research documents and sources. For the most prominent research documents, a co-citation network analysis based on references was executed to measure the mutual strength of the citations (see Figure 10). Among top references, Pang et al. (2002) have been cited by most of the documents. This article was the first empirical research on sentiment analysis of movie reviews through three different machine learning models. Researchers have widely cited this article because this article propounded the concept of sentiment analysis. Pang and Lee (2008) related to sentiment analysis and opinion mining have been cited multiple times by 107 other documents. Among the top references, Cambria (2016) has the least citation links among the top references. Erik Cambria in Cambria (2016) has cited the top two references only once in the same article.

Co-occurrence Analysis Based on Keywords

Keyword co-occurrence analysis provides the network of mutually employed keywords in terms of total link strengths. The bubble size of keywords shows the strength of co-occurrence and links with other keywords (see Figure 11). The network expresses that the word sentiment analysis occurred with machine learning, opinion mining, text mining (in red bubbles) and deep learning and natural language processing (in blue circle). This co-occurrence presents that the authors published the research on the theme of

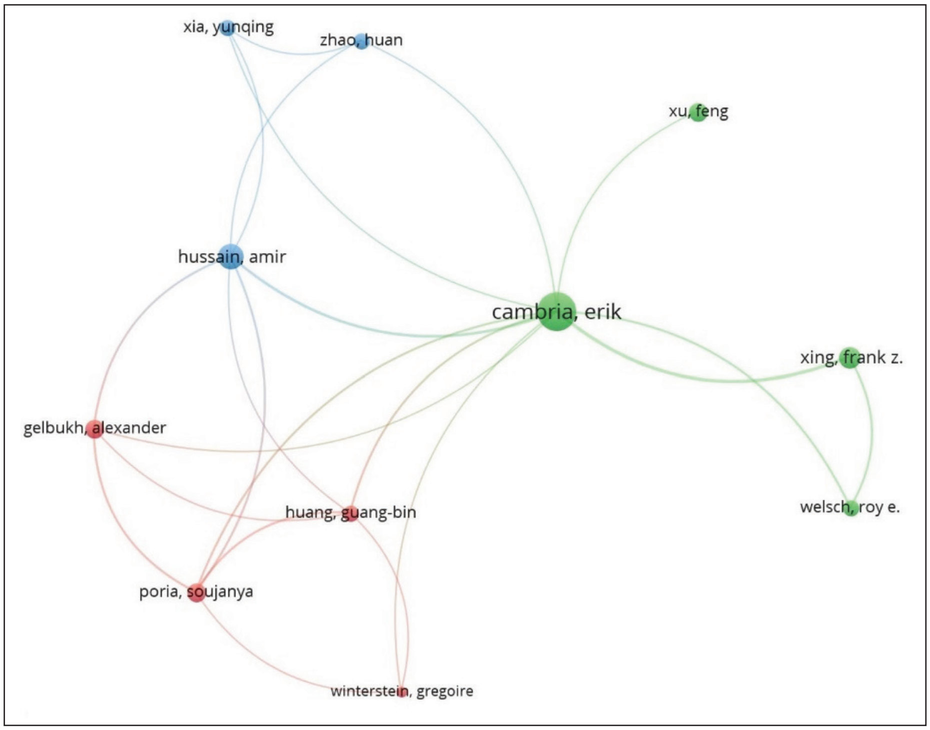

Figure 9. Co-authorship Analysis of the Most Productive Authors.

Source: Vosviewer.

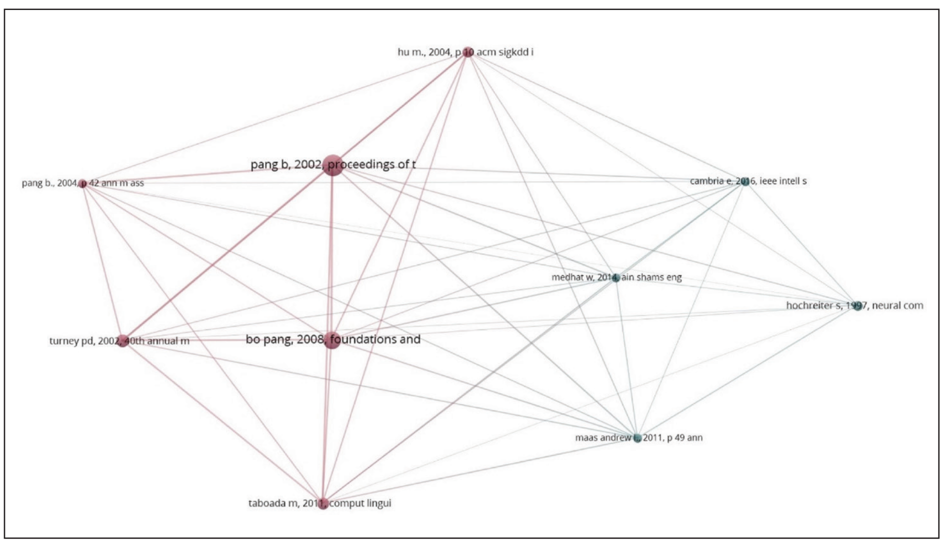

Figure 10. Co-citation Network Analysis Based on References for the Top Research Documents.

Source: Vosviewer.

sentiment analysis in the expertize of opinion and content analysis using advanced computing techniques. The corroborated themes are identified through the same colour bubbles.

The themes in the red circle present the use of sentiment analysis and machine learning for text analysis. The blue colour network depicts the co-occurrence of advanced computing techniques, that is natural language processing (NLP), long short-term memory, affective computing and big data models. Additionally, the yellow colour network shows the emerging topics including social media, Twitter and COVID-19. These networks state that big data analytics, text analysis for financial forecasting and social media sentiment analysis are the evolutionary and trending themes for research.

Thematic Evolution

In the form of a time series analysis of themes, a thematic evolution map was constructed for the period 2008–2022. Figure 12 exemplifies the evolutionary map for the theme

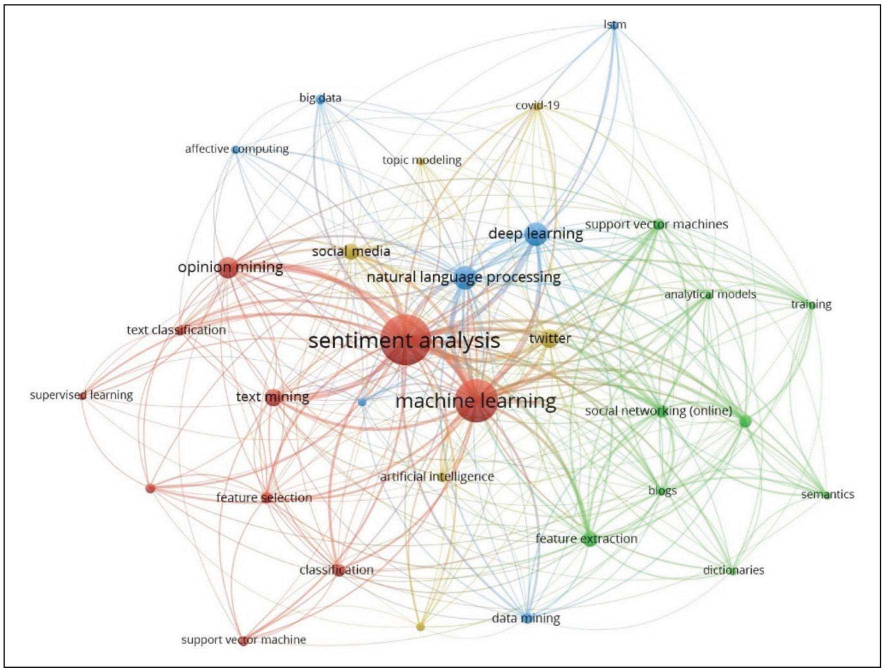

Figure 11. Co-occurrence Network Based on the Most Frequent Author Keywords.

Source: Vosviewer.

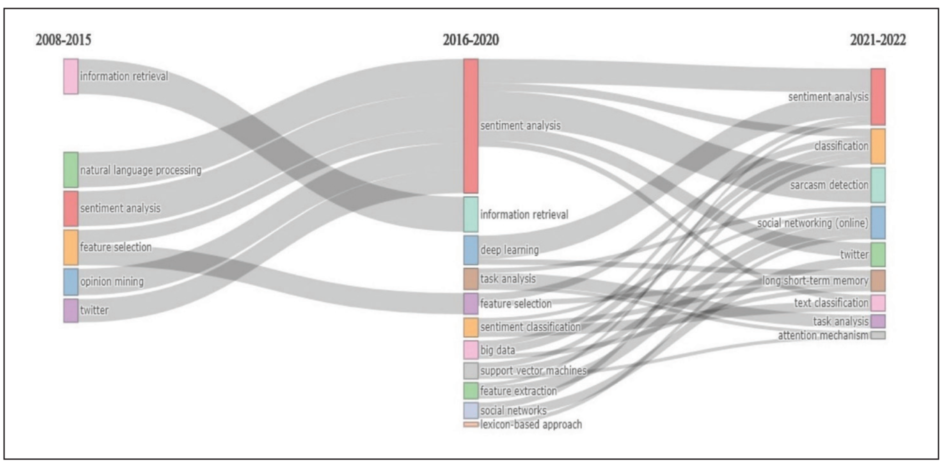

Figure 12. Thematic Evolution of Themes from 2008 to 2022.

Source: Biblioshiny.

of sentiment analysis and machine learning techniques along with the stock market prediction stream. In the period of 2008–2015, most of the research were conducted related to information retrieval and opinion mining using some traditional machine learning models, that is natural language processing. In the 2016–2020 period, other types of studies also conducted along with filtration of information. Gradually and consistently, other themes emerged during 2016–2020. For instance, deep learning, task analysis, sentiment classification, big data, support vector machines, feature extraction, social networks and lexicon-based approach evolved during this period along with the core theme of sentiment analysis. These themes relate to advanced techniques and networks along with the major sentiment analysis concept. This stream highly merged with sarcasm detection and Twitter in the period 2021–2022. With the advancement of technology and social networking platforms, the sentiment analysis concept is developed with social media platforms. Along with this concept, the theme classification evolved from previous period themes, that is feature selection, social networks, big data and feature selection along with the major theme of sentiment analysis.

Conceptual Evaluation of Themes Through a Structural Map: Structural Map and Dendrogram Plot Analysis

Bibliographic clustering based on mutual collaboration, future scope of themes and extent of research was executed in the analysed research domain. Conceptual structural analysis was performed by implementing the through multiple correspondence analysis and dendrogram plot (Figures 13 and 14, respectively). The conceptual structure graph is the two-dimensional plot that depicts the extent and future scope of research on the *y*-axis and *x*-axis, respectively. Based on these dimensions, overall themes constituted in the corpus of 610 research documents are segmented into major 4 clusters. The density of theme clusters presents the saturation of research themes. Clusters with a high density of points are highly researched areas having the potential for future scope. Contrary to this, scant clusters are the areas that have the emergence in the future. Dendrogram plot generates a hierarchal tree of themes that shows closure and similarity among theme clusters through U-shaped figures. The height of U-shaped lines between adjacent themes reveals the association and gap.

Initially, the overall research themes are segregated into four clusters that are as follows:

Cluster 1: Sentiment analysis and predictions using machine learning

Research studies referring the theme of sentiment analysis, machine learning and prediction fall in the densest cluster coloured red (in Figure 13) and violet (in Figure 14). As per the dimensions of the graph, this theme group is highly saturated which shows that a huge set of research publications related to this theme have published in various renowned journals. Additionally, most of the inter-connected themes have future emergence and scope. These research studies need more advancement in the form of data analytics and machine learning models in the future. About 363 research publications that are associated with these themes fall in this cluster. For instance, Ravi and Ravi (2015) and Tang et al. (2015) presented the application of machine learning, deep learning and natural language techniques for sentiment analysis that projects an emerging theme in the days to come. Sreesurya et al. (2020) applied the role of sentiment analysis using business intelligence tools for business effectiveness. Bouktif et al. (2020) predicted the stock market movements through technical data analytics and sentiment polarity. Research cluster points, that is neural networks, algorithms, models, networks and data extractions for mining are the prominent areas that will remain in the emergence with varied application domains.

Cluster 2: Social-media polarity detection

Studies in the field of social media polarity detection are shown in blue (Figure 13) and green colour (Figure 14). These research themes are reflected as under-researched due to less saturated research publications and continuous identification of closely related themes for future research. As per the dimensions, cluster points show immense sweep for the future due to the advancement of social media networks and the prominence of behaviour diversity in different fields. Themes such as social media networking platforms and Twitter-based polarity detection are highly under-researched which shows that these topics are novice in the research field and coming research studies will be performed on these topics. Publications fall in this cluster applied the polarity detection research in several ways, for instance, fake news detection (Kaliyar et al., 2020), in music album success forecasting (Cosimato et al., 2019) and polarity generation of words (Vechtomova, 2017).

Cluster 3: Feature classification and text analysis techniques

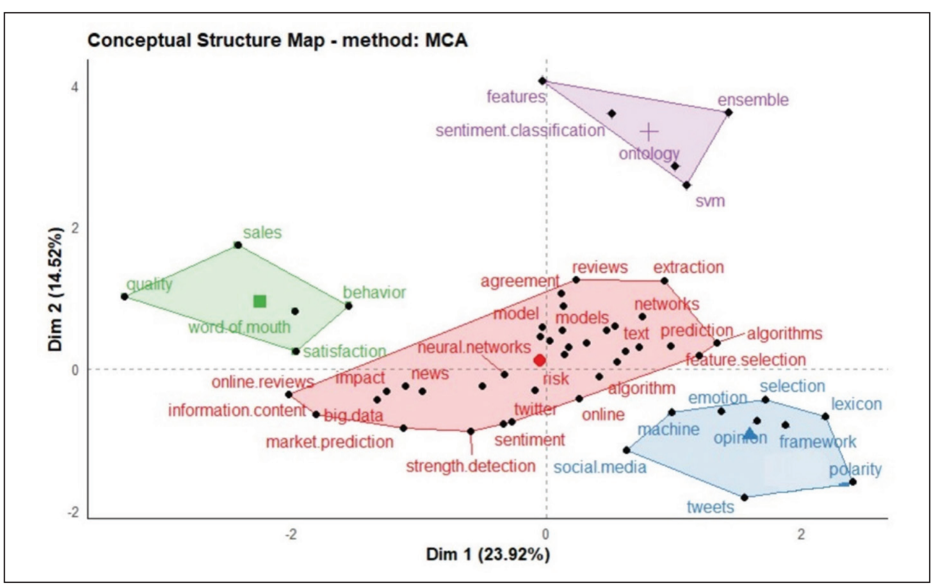

Figure 13. Conceptual Structure Map Based on Four Clusters.

Source: Biblioshiny.

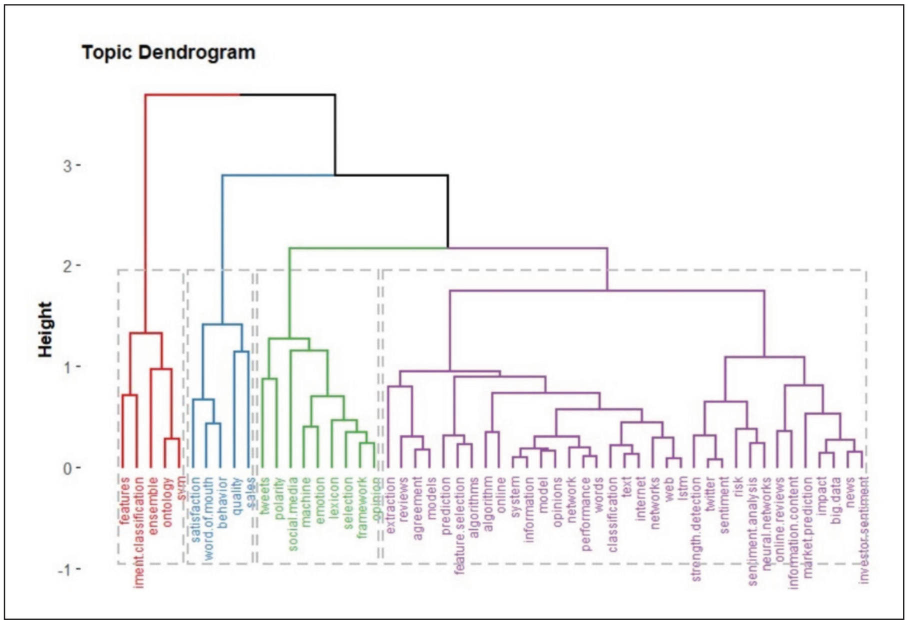

Figure 14. Dendrogram Plot Graph Based on Four Clusters.

Source: Biblioshiny.

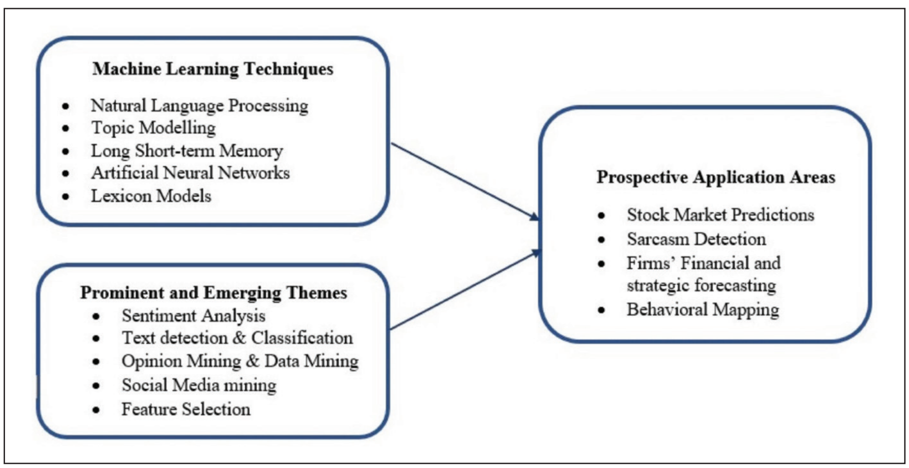

Figure 15. Framework for Emerging Research Themes.

Research themes in the third largest cluster are highlighted in violet (Figure 13) and red (Figure 14). The graph dimensions present that the cluster is placed at a location where the extent of research and future scope is high. It reflects that the research topics which fall in this category are highly adopted by the researchers for studies. Additionally, these have huge potential for studies in the future using embedded themes along with core areas. Cluster data points depict that the use of SVM is less than the ensemble models for text analysis and feature classifications.

Cluster 4: Behaviour, satisfaction and sales

Publications and researches that investigate behaviour and sales aspects are found to be outdated due to negligible scope in the research in the present era (see green cluster in Figure 13 and blue in Figure 14). The studies are limited in terms of analysing the behavioural and satisfaction aspects for sales using opinion mining and text classification.

Since Cluster 1 is dense and contains highly researched and identified areas which would continue to develop in the future. Furthermore, Clusters 2 and 3 are the second and third largest areas with under-researched and highly researched themes, respectively. On the other hand, Cluster 4 is a sparse area of non-trending research topics that have no scope in the future.

Discussions

This section presents the major findings of the bibliometric analysis and emerging areas with prospective research sweep as an answer all research questions proposed in the inception.

Major Findings of the Research Study

In the form of an answer to the first question, a descriptive analysis was conducted which presents the findings in the below points.

India and China are the countries where the researchers have largely contributed their efforts in the field of sentiment analysis. During a period of 15 years, these countries emerged as a hub of research in the field of sentiment analysis through various machine learning techniques. These studies are also cited and used by other countries' researchers for applying in different financial domains. In the context of TCs, India, China and the United Kingdom are the most eminent countries, whereas Belgium is the most cited country in term of C/P (see Table 2). In terms of country collaboration, Pakistan and Saudi Arabia are found as the most cooperative and productive countries among the top 10 country partners (see Figure 4). In the form of the most productive and most cited source, *IEEE Access* and *Expert Systems with Applications* are found as the top-ranked among all journals which are specialised in artificial intelligence, machine learning and other computing techniques (see Table 3 and Figure 5). In terms of most cited and productive authors in the analysed study expertise, Erik Cambia (Cambria, E.) is found as the top author who is a founder of SenticNet company and a prominent researcher with a focus on sentiment analysis, financial forecasting and dialogue systems (see Table 4). From Figure 5, sentiment analysis, classification, model, impact, Twitter, feature selection and social media are found to be the most prominent and emerging topics from the research.

As an answer to the second research question of the study, the following sub-points illustrate the key findings.

In terms of total link strength for bibliography and citations, *IEEE Access* (source), Erik Cambria (author) and India (country) are found to be at the top place among all sources, authors and countries, respectively (see Tables 5 and 6). From Figure 9, Erik Cambria is found as the most prominent author having the highest collaboration with other authors in the field of sentiment analysis and advanced machine learning models. The author is found in the cluster that has the highest links among all authors. In the form of mutual citations, Pang et al. (2002) have the maximum citation interlinks with other documents. Following it, Pang and Lee (2008) are placed as the second most interlinked and referred document (see Figure 10). The top author keywords having high co-occurrence and links are found as sentiment analysis, machine learning, natural language processing, opinion mining, Twitter, online social media networking, feature selection, task analysis and social media (Figure 11). From the time series thematic evolution, sentiment analysis is found as the most prominent theme which mainly emerged after 2016 with an application of natural language processing, opinion mining and feature selection techniques. During 2021–2022, the theme shows an emergence with advanced themes, that is social media networking, long- short-term memory model, text classification and classification (see Figure 12). From Figure 13, the conceptual theme cluster 'sentiment analysis, machine learning, and prediction' is found as the most saturated research area with high research publications (red cluster). On the other hand, the second densest cluster 'social media and polarity detection' is identified as the research area with a high future sweep for research (blue cluster).

In the form of the third answer of this study, the conceptual framework for potential and emerging themes was developed through the findings of co-occurrence network, thematic map and conceptual structure plot. The framework shows some prosperous areas where future researchers can implement the core sentiment analysis and related concepts.

Figure 15 illustrates the conceptual framework for the core theme of sentiment analysis and related sub-themes. The framework shows that the concept of machine learning techniques-driven sentiment analysis, text and opinion mining can contribute to various financial and strategic areas, that is, stock market predictions, Sarcasm detection, strategic predictions and behavioural mapping through analysis of opinions. Future research studies can combine these themes to produce the most effective and advanced results for researchers and organizations.

Theoretical and Practical Impact

This study will add to the body of knowledge and will help the researchers to identify and adopt the emerging themes related to machine learning-based sentiment analysis and stock market prediction. Along with thematic assistance, it provides knowledge to the researchers about the prominent countries where sentiment analysis-based studies are frequently held. This will enable them to collaborate with the authors of these highly prolific countries. Furthermore, the article shows the most effective sources which include a high volume of research in the domain of sentiment analysis with ML techniques.

Besides above, this research supports the firms operating in different fields of finance. It provides information about the application of sentiment analysis in different domains, that is fraud detection, opinion mining, social media networking, financial market predictions, text analysis, classification of views and polarity generation. Firms can implement ML and AI-based sentiment models for generating the opined views of stakeholders. Contribution in the form of sentiment analysis-based study can aid the companies to implement machine learning models for analysing the sentiment polarity for making portfolio management decisions.

Limitations of the Bibliometric Study

This bibliometric analysis study provides various insightful sets of information, but it also occupies some limitations that can be removed in future research. First, the data extraction was performed through the Web of Science database only. Second, a particular search query was employed to extract the corpus. Other keywords can be used in future studies for getting the synthesis of other related fields. Third, the data period was limited to fifteen years to get more detailed results. Potential researchers can employ different study periods to get findings in another way. Additionally, the study does not deal with systematic analysis of literature which can be a direction for the future.

Future Scope and Direction of the Theme

The potential scope of sentiment analysis through machine learning is presented in the below-mentioned points.

1. Behavioural Mapping: In the field of advanced technology, behavioural mapping is an emerging area of research that can be linked with machine learning-based sentiment analysis. Upcoming researches employ a wide range of AI-based models, where opinion mining and sentiment classification can be used as a base.
2. Financial and strategic prediction: Future prediction for stock market prices, loan default, firm revenue and services' success is a rising research area in finance and other domains. Sentiment analysis and polarity detection, the major tools for generating the stakeholders' emotions can be embedded with machine learning models to predict future financial factors.

Conclusion

This article presents a novel bibliometric analysis and network visualization research study associated with sentiment analysis through machine learning models in the stock market between 2008 and 2022. Using the most effective authors, countries, journals and themes, the study illustrates that China is the most prominent country around the world that produced 223 research articles in the field of advanced technique-based sentiment classification. Moreover, the authors from China have largely collaborated with other countries' researchers. In addition to it, India is the most citation-driven country for the research theme. Furthermore, the prominent topic 'sentiment analysis and predictions using machine learning' reflects the future emergence with existing growth. This study adds value to the researchers as they can identify the countries and authors which are highly working on sentiment analysis theme. Additionally, it provides the potential usage of the sentiment analysis in stock market and financial predictions, sarcasm detection and behavioural mapping. Future researchers can embed these domains with sentiment analysis for providing support to firms' strategic decisions.

Acknowledgements

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article. The author reports a relationship with Kurukshetra University that includes non-financial support.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Shivani https://orcid.org/0000-0003-4111-8104

References

Agarwal, S., Kumar, S., & Goel, U. (2021). Social media and the stock markets: An emerging market perspective. *Journal of Business Economics and Management, 22*(6), 1614–1632. https://doi.org/10.3846/jbem.2021.15619

Ahlgren, O. (2016). Research on sentiment analysis: The first decade. In *2016 IEEE 16th International Conference on Data Mining Workshops* (pp. 890–899). https://doi.org/10.1109/ICDMW.2016.94

Ajitha, P., Sivasangari, A., Immanuel Rajkumar, R., & Poonguzhali, S. (2021). Design of text sentiment analysis tool using feature extraction based on fusing machine learning algorithms. *Journal of Intelligent and Fuzzy Systems*, *40*(4), 6375–6383. https://doi.org/10.3233/JIFS-189478

Al-Rubaiee, H., Qiu, R., & Li, D. (2016). Analysis of the relationship between Saudi twitter posts and the Saudi stock market. In *2015 IEEE 7th International Conference on Intelligent Computing and Information Systems, ICICIS 2015*, (pp. 660–665). https://doi.org/10.1109/IntelCIS.2015.7397193

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

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

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, *31*(2), 102–107. https://doi.org/10.1109/MIS.2016.31

Carosia, A. E. O., Coelho, G. P., & Silva, A. E. A. (2020). Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media. *Applied Artificial Intelligence, 34*(1), 1–19. https://doi.org/10.1080/08839514.2019.1673037

Casoto, P., Dattolo, A., & Tasso, C. (2008). Sentiment classification for the Italian language: A case study on movie reviews. *Journal of Internet Technology, 9*(4), 365–373.

Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing*, *50*, 135–141. https://doi.org/10.1016/j.asoc.2016.11.022

Christina, N. (2021). *Stock market prediction using sentiment analysis*. International Hellenic University.

Cobo, M. J. (2020). Research trends in Sentiment Analysis and Opinion Mining from Knowledge Management approach : A science mapping from 2007 to 2020. In *2020 international conference on innovation and intelligence for informatics, computing and technologies* (pp. 1–6).

Cosimato, A., De Prisco, R., Guarino, A., Malandrino, D., Lettieri, N., Sorrentino, G., & Zaccagnino, R. (2019). The conundrum of success in music: Playing it or talking about it? *IEEE Access*, *7*, 123289–123298. https://doi.org/10.1109/ACCESS.2019.2937743

Edmans, A., Garcia, D., & Norli, Ø. (2007). Sports sentiment and stock returns. *The Jounral of Finance*, *62*(4), 1967–1998.

Feldman, R. (2013). Techniques and applications for sentiment analysis: The main applications and challenges of one of the hottest research areas in computer science. *Communications of the ACM*, *56*(4), 82–89. https://doi.org/10.1145/2436256.2436274

Garcia-Lopez, F. J., Batyrshin, I., & Gelbukh, A. (2018). Analysis of relationships between tweets and stock market trends. *Journal of Intelligent and Fuzzy Systems*, *34*(5), 3337–3347. https://doi.org/10.3233/JIFS-169515

Grljević, O., Bošnjak, Z., & Kovačević, A. (2022). Opinion mining in higher education: A corpus-based approach. *Enterprise Information Systems*, *16*(5), 1–26. https://doi.org/10.1080/17517575.2020.1773542

Hassanein, A., Mostafa, M. M., Benameur, K. B., & Al-Khasawneh, J. A. (2021). How do big markets react to investors' sentiments on firm tweets? *Journal of Sustainable Finance & Investment*, 1–23. https://doi.org/10.1080/20430795.2021.1949198

Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet—A deep convolutional neural network for fake news detection. *Cognitive Systems Research*, *61*, 32–44. https://doi.org/10.1016/j.cogsys.2019.12.005

Keramatfar, A., & Amirkhani, H. (2019). Bibliometrics of sentiment analysis literature. *Journal of Information Science*, *45*(1), 3–15. https://doi.org/10.1177/0165551518761013

Kinyua, J. D., Mutigwe, C., Cushing, D. J., & Poggi, M. (2021). An analysis of the impact of President Trump's tweets on the DJIA and S&P 500 using machine learning and sentiment analysis. *Journal of Behavioral and Experimental Finance*, *29*, 1–14. https://doi.org/10.1016/j.jbef.2020.100447

Krishnamoorthy, S. (2018). Sentiment analysis of financial news articles using performance indicators. *Knowledge and Information Systems, 56*(2), 373–394. https://doi.org/10.1007/s10115-017-1134-1

Leitch, D., & Sherif, M. (2017). Twitter mood, CEO succession announcements and stock returns. *Journal of Computational Science*, *21*. https://doi.org/10.1016/j.jocs.2017.04.002

Li, Xiao, Shen, D., Xue, M., & Zhang, W. (2017). Daily happiness and stock returns: The case of Chinese company listed in the United States. *Economic Modelling*, *64*, 496–501. https://doi.org/10.1016/j.econmod.2017.03.002

Li, Xin, & Lei, L. (2021). A bibliometric analysis of topic modelling studies (2000–2017). *Journal of Information Science, 47*(2), 161–175. https://doi.org/10.1177/0165551519877049

Lin, Y., Li, J., Yang, L., Xu, K., & Lin, H. (2020). Sentiment analysis with comparison enhanced deep neural network. *IEEE Access, 8*, 78378–78384. https://doi.org/10.1109/ACCESS.2020.2989424

Liu, B. (2012). Sentiment analysis: A fascinating problem. In *Sentiment analysis and opinion mining* (pp. 1–8). Springer International Publishing. https://doi.org/10.1007/978-3-031-02145-9_1

Lopez-Martinez, R. E., & Sierra, G. (2021). State of research on natural language processing in Mexico—A bibliometric study. *Journal of Data, Information and Management*, *3*(3), 183–195. https://doi.org/10.1007/s42488-021-00051-5

Mcgurk, Z., Nowak, A., & Hall, J. C. (2020). Stock returns and investor sentiment: Textual analysis and social media. *Journal of Economics and Finance, 44*, 458–485.

Mendoza-Urdiales, R. A., Núñez-Mora, J. A., Santillán-Salgado, R. J., & Valencia-Herrera, H. (2022). Twitter sentiment analysis and influence on stock performance using transfer entropy and EGARCH methods. *Entropy, 24*(7), 874. https://doi.org/10.3390/e24070874

Mohamed, A. (2017). An evaluation of sentiment analysis and classification algorithms for Arabic textual data. *International Journal of Computer Applications*, *158*(3), 29–36. https://doi.org/10.5120/ijca2017912770

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*, 125–144. https://doi.org/10.1016/j.eswa.2016.12.036

Oneto, L., Bisio, F., Cambria, E., & Anguita, D. (2017). Semi-supervised learning for affective common-sense reasoning. *Cognitive Computation*, *9*(1), 18–42. https://doi.org/10.1007/s12559-016-9433-5

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)* (pp. 1345–1350). https://doi.org/10.1109/SCOPES.2016.7955659

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *1*(2), 91–233. https://doi.org/10.1561/1500000001

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up ? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 conference on empirical methods in neural language processing (EMNLP 2002)* (pp. 79–86). https://doi.org/10.3115/1118693.1118704

Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, *135*, 60–70. https://doi.org/10.1016/j.eswa.2019.06.014

Poria, S., Gelbukh, A., Cambria, E., Hussain, A., & Huang, G. Bin. (2014). EmoSenticSpace: A novel framework for affective common-sense reasoning. *Knowledge-Based Systems, 69*(1), 108–123. https://doi.org/10.1016/j.knosys.2014.06.011

Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis. In *2012 IEEE/ACM international conference on advances in social networks analysis and mining analyzing* (pp. 119–123). https://doi.org/10.1109/ASONAM.2012.30

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, *89*(June). https://doi.org/10.1016/j.knosys.2015.06.015

Reboredo, J. C., & Ugolini, A. (2018). The impact of Twitter sentiment on renewable energy stocks. *Energy Economics, 76*, 153–169. https://doi.org/10.1016/j.eneco.2018.10.014

Sakhare, N. N., Imambi, S. S., Kagad, S., Kapadwanjwala, T., Malekar, H., & Dalal, M. (2020). Stock market prediction using sentiment analysis. *International Journal of Advanced Science and Technology*, *29*(4, Special Issue).

Saurabh, S., & Dey, K. (2020). Unraveling the relationship between social moods and the stock market: Evidence from the United Kingdom. *Journal of Behavioral and Experimental Finance*, *26*, 1–9. https://doi.org/10.1016/j.jbef.2020.100300

Shi, D., He, G., Cao, S., Pan, W., Zhang, H.-Z, Yu, D., & Hung, M.-C. (1992). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Molecular Carcinogenesis*, *5*(3), 213–218. https://doi.org/10.1002/mc.2940050308

Singh, L. G., & Singh, S. R. (2021). Empirical study of sentiment analysis tools and techniques on societal topics. *Journal of Intelligent Information Systems*, *56*(2), 379–407. https://doi.org/10.1007/s10844-020-00616-7

Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2013). Predictive sentiment analysis of tweets: A stock market application. In *International workshop on human-computer interaction and knowledge discovery in complex, unstructured, big data* (pp. 77–88). Springer. http://link.springer.com/10.1007/978-3-642-39146-0_8

Sreesurya, I., Rathi, H., Jain, P., & Jain, T. K. (2020). Hypex: A tool for extracting business intelligence from sentiment analysis using enhanced LSTM. *Multimedia Tools and Applications*, *79*(47–48), 35641–35663. https://doi.org/10.1007/s11042-020-08930-6

Sul, H. K., Dennis, A. R., & Yuan, L. I. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, *48*(3), 454–488. https://doi.org/10.1111/deci.12229

Tabari, N., Seyeditabari, A., Peddi, T., Hadzikadic, M., & Zadrozny, W. (2019). A comparison of neural network methods for accurate sentiment analysis of stock market tweets. In *ECML PKDD 2018 workshops* (Vol. 11054, pp. 51–65). Springer International Publishing. https://doi.org/10.1007/978-3-030-13463-1_4

Tan, S. D., & Tas, O. (2021). Social media sentiments in international stock returns and trading activity. *Jounral of Behavioral Finance*, *22*(2), 221–234.

Tang, D., Qin, B., & Liu, T. (2015). Deep learning for sentiment analysis: Successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *5*(6), 292–303. https://doi.org/10.1002/widm.1171

Teti, E., Dallocchio, M., & Aniasi, A. (2019). The relationship between twitter and stock prices. Evidence from the US technology industry. *Technological Forecasting and Social Change, 149*(March), 1–9. https://doi.org/10.1016/j.techfore.2019.119747

Uhl, M. W., & Ag, U. (2014). Reuters sentiment and stock returns. *Journal of Behavioral Finance*, *15*(February 2015), 287–298. https://doi.org/10.1080/15427560.2014.967852

Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2022). Does Twitter affect stock market decisions? Financial sentiment analysis during pandemics: A comparative study of the H1N1 and the COVID-19 periods. *Cognitive Computation*, *14*, 372–387. https://doi.org/10.1007/s12559-021-09819-8

van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, *111*(2), 1053–1070. https://doi.org/10.1007/s11192-017-2300-7

Vechtomova, O. (2017). Disambiguating context-dependent polarity of words: An information retrieval approach. *Information Processing and Management*, *53*(5), 1062–1079. https://doi.org/10.1016/j.ipm.2017.03.007

Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing, 36*(3), 492–508. https://doi.org/10.1016/j.ijresmar.2019.01.010

Vijay, N., Singh, S., & Malhotra, G. (2018). Sentiment analysis: Gauging the effect of news on stock prices in Indian Stock Market. *International Journal of Trade, Economics and Finance*, *9*(4), 148–152. https://doi.org/10.18178/ijtef.2018.9.4.605

Xia, R., Xu, F., Yu, J., Qi, Y., & Cambria, E. (2016). Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. *Information Processing and Management, 52*(1), 36–45. https://doi.org/10.1016/j.ipm.2015.04.003

Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Intelligent asset allocation via market sentiment views. *IEEE Computational Intelligence Magazine, 13*(4), 25–34. https://doi.org/10.1109/MCI.2018.2866727

Xu, K., Pang, Y., & Han, J. (2022). Dynamic cross-correlation between online sentiment and stock market performance: A global view. *Discrete Dynamics in Nature and Society, 2021,* 1–11. https://doi.org/10.1155/2021/6674379

Yadav, A., Jha, C. K., Sharan, A., & Vaish, V. (2020). Sentiment analysis of financial news using unsupervised approach.

Procedia Computer Science, *167*(2019), 589–598. https://doi.org/10.1016/j.procs.2020.03.325

Yang, S. Y., Mo, S. Y. K., & Liu, A. (2015). Twitter financial community sentiment and its predictive relationship to stock market movement. *Quantitative Finance*, *15*(10), 1637–1656. https://doi.org/10.1080/14697688.2015.1071078

Yılmaz, B. (2022). *Market sentiment analysis: How it works & 7 data sources*. AI Multiple.

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter 'I hope it is not as bad as I fear'. *Procedia—Social and Behavioral Sciences, 26*(2007), 55–62. https://doi.org/10.1016/j.sbspro.2011.10.562

Zuopeng, J., Ranjan, P., Sharma, D., & Eachempati, P. (2021). Big data analytics and machine learning: A retrospective overview and bibliometric analysis. *Expert Systems with Applications*, *184*(July), 1–18. https://doi.org/10.1016/j.eswa.2021.115561

About the Authors

Shivani (ssaini31998@gmail.com) is a PhD Research Scholar at University School of Management, Kurukshetra University. She has a specialization in financial management and stock market analysis. She has presented seven research papers in national and international conferences.

Bhag Singh Bodla (bsbkuk@gmail.com) is a Retired Professor at the University School of Management, Kurukshetra University. He is specialized in banking and financial management, stock market analysis, portfolio management and insurance management. He has published more than 140 research papers in indexed and peer-reviewed journals and books. He has also presented more than 100 papers in several international and national conferences.