International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Bibliometric Analysis of Stock Market Sentiments Using Machine Learning

Soni[1], Nishant Kumar[2]

[1]Research Scholar, Department of Business Administration, Lucknow University
[2]Associate Professor Department of Business Administration, Lucknow University

Abstract
Sentiment analysis has gained significant attention in financial markets for its potential to gauge market sentiment and predict stock price movements. This bibliometric analysis explores the landscape of existing research literature in sentiment analysis using machine-learning techniques in the stock market domain. for this purpose, I, extract 572 research articles pertinent to the chosen field. These articles were published in 368 journals between January 2012 to April 2024 and have been listed in the Scopus database. By systematically reviewing relevant literature, this study aims to identify key trends, research themes, influential authors, and publication outlets in the field. The paper employed a blend of bibliometric and network analysis techniques using "R" and VOSviewer, such as co-citation analysis, keyword co-occurrence analysis, and author co-citation analysis to uncover sentiment analysis's intellectual structure and evolution in stock market research. The findings of this study provide valuable insights for practitioners, researchers, and policymakers interested in understanding the advancements and future directions of research in sentiment analysis of investors in the stock market.

Keywords: Bibliometric Analysis, Sentiment Analysis, Stock Market, Financial Market, Machine Learning, and Natural Language Processing.

1. Introduction

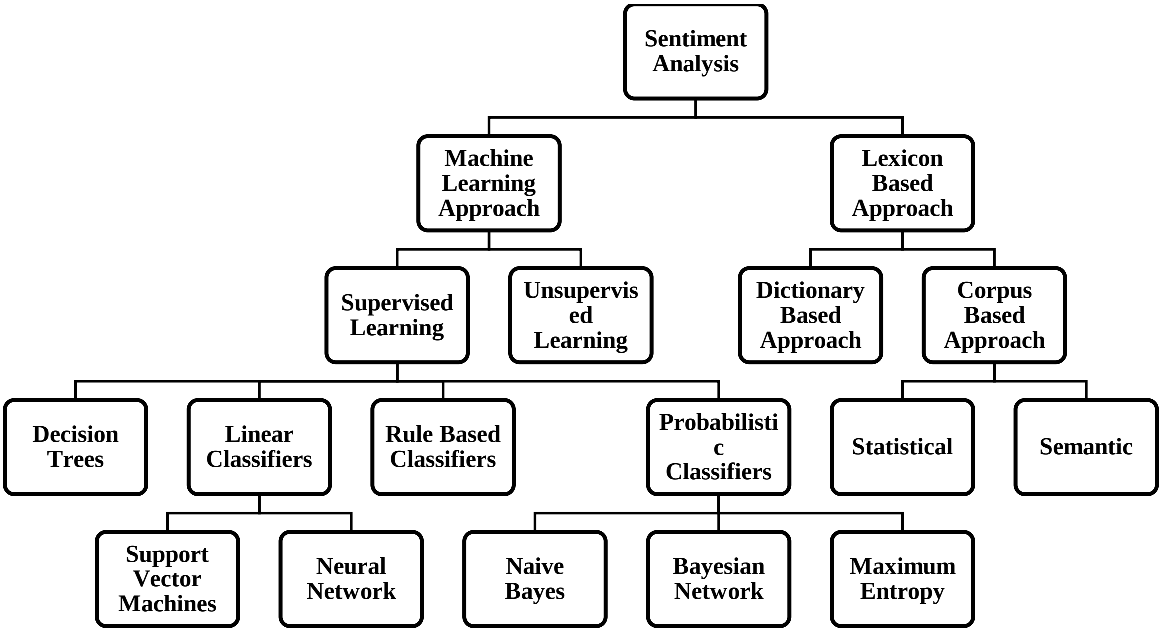
Sentiment analysis, a subfield of natural language processing (NLP), has gathered significant attention in recent years due to its potential applications in understanding market sentiments in various domains, including the stock market. In the context of the stock market, sentiment analysis can provide valuable insights into investor emotions, that can influence short-term market trends and stock prices. Understanding human behavior is paramount in the era of advanced computing and information technology. Amidst various behavioural biases, the study of sentiments and opinions has emerged as a pivotal tool for analysing the moods and beliefs of individuals in the twenty-first century. Liu (2012) defines 'sentiment' as a quintuple (comprising five elements: E, A, S, H, T), where 'E' represents an entity, 'A' denotes an aspect, 'S' signifies the sentiment towards entity 'E's aspect 'A', 'H' identifies the holder of the opinion, and 'T' indicates the time of opinion expression. Building upon this foundation, (Lee, 2002) introduced the concept of sentiment analysis primarily focusing on the classification of movie reviews. As elucidated by Duyu Tang (2015), Sentiment analysis refers to the techniques employed to extract polarity from individuals' emotions, moods, and opinions conveyed through textual data and other informational sets. Through sentiment classification and polarity generation, sentiments are categorized into positive, negative, or neutral, offering a nuanced viewpoint perspective (FELDMAN, 2013). during the Web 2.0

era, social networks have transformed communication. platforms like Facebook Twitter and other social networking sites are used for expressing opinions on various topics such as news, movies, events, and product reviews. This user-generated content is valuable for business analysts, who use this information to understand public opinion and improve their products and strategies. Sentiment analysis plays a vital role in extracting and classifying these opinions as positive, negative, or neutral (Bhardwaja, 2015).

Sentiment Classification Techniques.

Sentiment classification techniques can be broadly divided into two categories:

Figure 1:



In the realm of analyzing the stock market sentiments using machine learning, there exists a compelling need for bibliometric analysis. While numerous scholars have attempted to do so, several limitations persist in their approaches. Such as **Bodlal (2023)** talks about bibliometric analysis, concentrating on stock market prediction and utilizing the Web of Science (WOS) database as the primary data source. This study excludes consideration of alternative databases such as Scopus. Specific keywords employed were 'sentiment analysis' AND 'machine learning' or 'stock market prediction'**.** Using Scopus data, **Zulfikar(2022)** has also focused on stock market performance throughout the COVID-19 outbreak and takes data from 2020 till end of 2021 and has not done a comprehensive study and uses limited keywords. Therefore. the article covered a few studies and an overview was compromised**. Pooja Bagane et al..(2021)** have limited the descriptive style of the literature review and focused only on the LSTM model of sentiment analysis ignoring other sentiment analysis techniques and limiting the keywords to "stock market prediction" and "LSTM". **Henrique,(2019)** focused only on fifty-seven documents specifically related to the North American market and used a prediction model involving support vector machines (SVMs) and neural networks.

After a comprehensive review of the existing literature, it is evident that there exists a gap in prior researches which have undertaken bibliometric analysis of stock market sentiment using machine learning in not a comprehensive manner.The present study has made use of the Scopus database for data acquisition and has employed bibliometric analysis techniques utilizing "R" biblioshiny and VOSviewer software. Bibliometric analysis is a popular and rigorous method for analyzing large volumes of scientific data. The paper attempts an in-depth examination of bibliometric methodology as provided by **Naveen Donthu, (2021)** encompassing various techniques such as performance analysis, science mapping, network analysis, citation analysis, co-citation analysis, and bibliography coupling.

Objectives of the study

The main objective of this study is to present state of the art research on stock market sentiments using machine learning by answering the following five research questions (RQ):

1. How has the publication trend of sentiment analysis in the stock market using machine learning evolved, as indicated by the number of papers published annually?
2. What are the primary publication outlets for influential works on sentiment analysis in the stock market using machine learning, such as prominent research papers and articles in the reputed journals?
3. Who are the most prolific contributors to the research on sentiment analysis in the stock market using machine learning and which countries and organizations do they belong to?
4. What insights can previous studies provide about sentiment analysis in the stock market using machine learning about prevalent themes explored in the literature?
5. Where are the potential avenues for further research in advancing the understanding of sentiment analysis in stock market using machine learning?

2. Literature Review

In recent years, there has been a surge in studies focusing on stock market sentiment using machine learning techniques**. Gao's** study **(2010)** implements an approach involving the N-gram model for Chinese news word detection and integrates Appraisal theory into sentiment analysis. Classification algorithms including Naïve Bayes, K-nearest Neighbor, and Support Vector Machine are employed for sentiment classification. **Porshnev(2013)** developed linguistic technologies that investigated users' moods and psychological states of people. Further, he used the lexicon-based approach to analyze the psychological state. **Li Im, (2013)** in his paper, analyses business and financial news to analyze the sentiment using a lexicon-based approach. **Pagolu,(2016)** concluded that Twitter was the most suitable for researchers to study public sentiment. **Dash,(2023)** highlighted the potential of machine learning in forecasting Indian stock trends, particularly when extracting sentiment from financial data. **Chiong (2018)** proposed a sentiment analysis-based approach for financial market prediction, which outperformed a deep learning model in terms of time and accuracy**. Seals (2020)** investigated sentiment analysis in stock forecasting, finding that it could improve performance when combined with historical data. **Das et al. (2024)** introduce a multitasking sequence-to-sequence model integrating sentiment and emotion analysis with financial investment analysis, achieving high accuracies in sentiment identification and emotion recognition. **Patil et al. (2022)** present research on stock trend prediction using natural language processing, demonstrating the efficacy of machine learning models like LSTM in sentiment analysis and market trend correlation. Also, they perform experiments on real-world datasets using machine learning models like support vector machine (SVM), Random forest classifier, and Decision tree classifier followed by deep learning models like LSTM, RNN, and CNN neural networks. **Gaurav and Kotrappa (2020)** propose sentiment-aware stock forecasting models utilizing Log BiLinear (LBL) and Recurrent Neural Network (RNN) architectures, showcasing improved performance over traditional deep learning models. **Chong and Shah (2022)** compare the performance of Naive Bayes and SVM classifiers in sentiment analysis of healthcare companies' stock comments in Bursa Malaysia, emphasizing the significance of hyperparameter tuning for optimization. **Alazba et al. (2020**) conducted sentiment analysis on Arabic tweets related to the Saudi stock market, utilizing machine learning classifiers to achieve high accuracy in sentiment classification. **Sangsavate et al. (2019)** compare the performance of Naïve Bayes and SVM classifiers in sentiment classification of Thai FinTech news, highlighting SVM's superior performance. **Rajput and Dubey (2017)** discuss sentiment analysis in the stock market, emphasizing the importance of supervised learning methods like Naïve Bayes and SVM in categorizing public opinions effectively. **Lin et al. (2014)** integrate web news media sentiment analysis into stock trading signal prediction, demonstrating the effectiveness of sentiment indicators in enhancing prediction accuracy and profit generation.

Overall these studies contribute to advancing sentiment analysis techniques and their application in stock market forecasting, highlighting the potential for improved trading strategies and decision-making processes. These studies collectively underscore the potential of machine learning in stock market sentiment analysis highlighting the role of machine learning based methodology in prospective studies as well.

3. Methodology

3.1 Bibliometric Search

The bibliometric search for publications related to this review paper focuses on data from 2012 to 2024 **(13 years)**. The search process involved a six-stage method which includes database extraction, year filtrations, subject filtration, document filtration, keyword filtration, and language filtration, **(refer to Figure 2).**

Stage 1: For the bibliometric analysis of stock market sentiments using machine learning, the initial stage involved database search. Scopus database is popularly known for its comprehensive coverage of bibliometric information and its stringent indexing criteria for scientifically and intellectually relevant publications. Thus, Scopus is widely recommended for bibliometric evaluations and is particularly useful for compiling large corpora for review. It is considered a high-quality source for bibliometric data, with a strong correlation of its measures with those from other scientific databases like Web of Science etc.

The search keywords ***((‘Stock Market’ OR ‘Financial Market’) AND (‘Natural Language Processing’ OR ‘Machine Learning’) AND (‘Market Sentiment’))*** were selected to align with the central theme of the research. The search was limited to articles published from 2012 to 2024, yielding a total of 946 articles in the database.

Stage 2: Year filtration is the second stage where we chose to include data from 2012 to 2024 (13 years).As the number of articles on the use of machine learning in finance upto the year 2011 were almost negligible, the year 2011 was filtered and 940 articles got selected.

Stage 3: The third stage is ‘Subject’ filtration which led us to include publications solely from (1) economics, econometrics, and finance, (2) business management and accounting, (3) Computer Science**,** and (4) Engineering.. The subject filtration eliminated 65 items, leaving 875 to be considered for analysis.

Stage 4: The fourth stage is ‘document’ filtration. Only source type list included here are journal articles and conference proceedings since they are typically evaluated based on originality and subjected to rigorous peer review, both being necessary criteria. Other sources, such as books and book chapters, were excluded since they typically failed to meet these criteria. This screening eliminated 164 articles, leaving only 711 articles.

Stage 5: The fifth stage is keyword filtration. At this stage, we select the most relevant keywords,i.e., sentiment analysis, neural language processing, opinion mining, stock market, machine learning, etc. Some other keywords were also excluded that were not directly related to our topic. The keyword filtration eliminated 127 articles, leaving only 584 articles.

Stage 6: The sixth stage is Language filtration. Items authored in English only were included for bibliometric review. Thus, only 572 items were finally included after language filtration, eliminating 12 more articles.

A total no. of 946 articles were filtered out considering year, document, keywords, language, and subject criteria. After the entire filtration process, 572 articles were retained for bibliometric analysis.

3.2 Bibliometric analysis

In bibliometric analysis, a literature review on stock market sentiments was conducted employing a Machine learning approach. A total of 572 articles sourced from the Scopus database formed the basis of this study. Various bibliometric analyses were performed to outline publication trends, identify prominent contributors (authors, countries, institutions), Publication performance (local citation, global citation) publications sources (journals, articles), and prevalent keywords. VOSviewer has also been utilized to enhance the exploration of key contributors and publications using keywords co-occurrence analysis, Countries' Co-authorship analysis, and Bibliographic Coupling.

Figure 2: Bibliometric Data Extraction: a step-by-step procedure

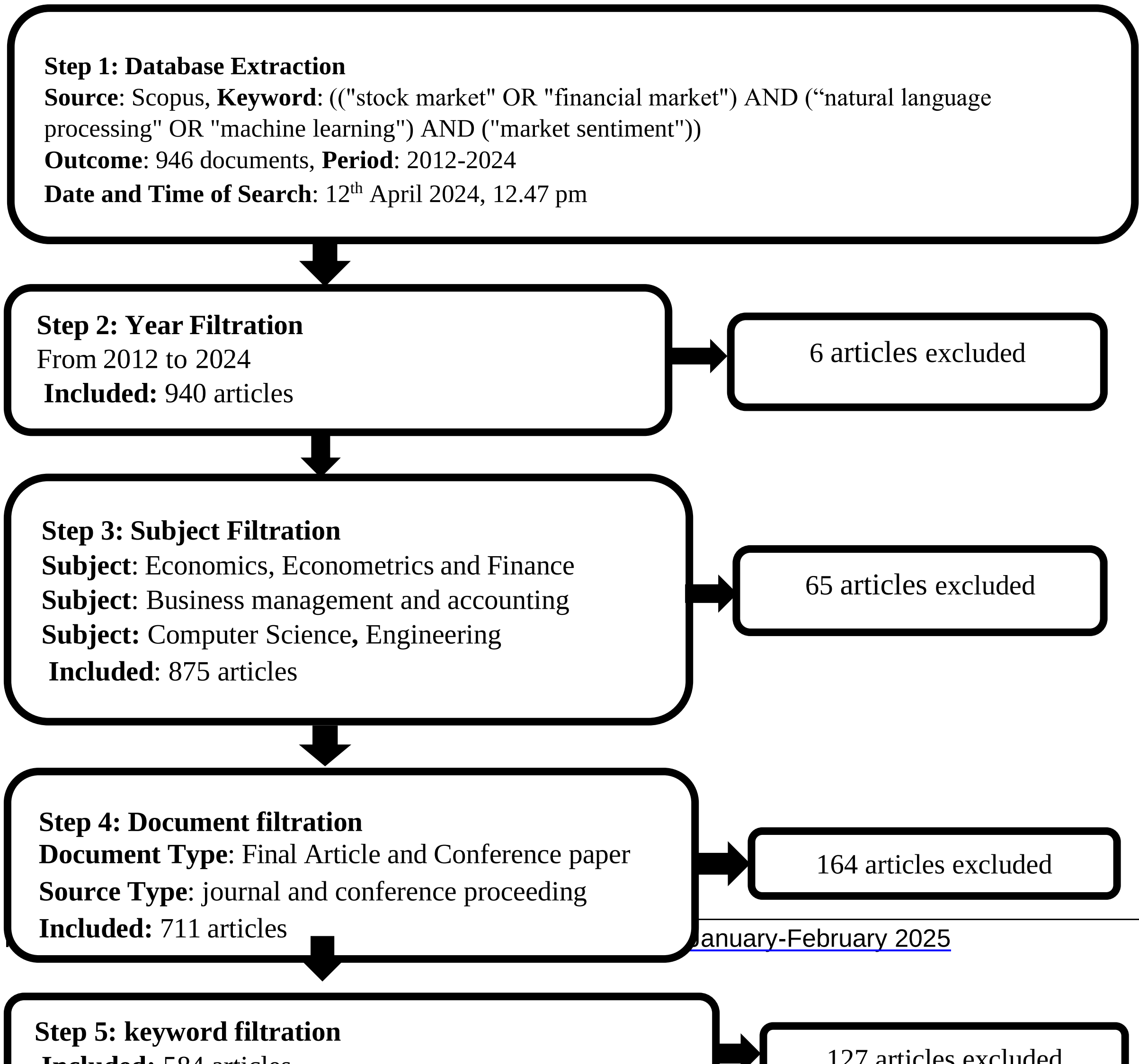
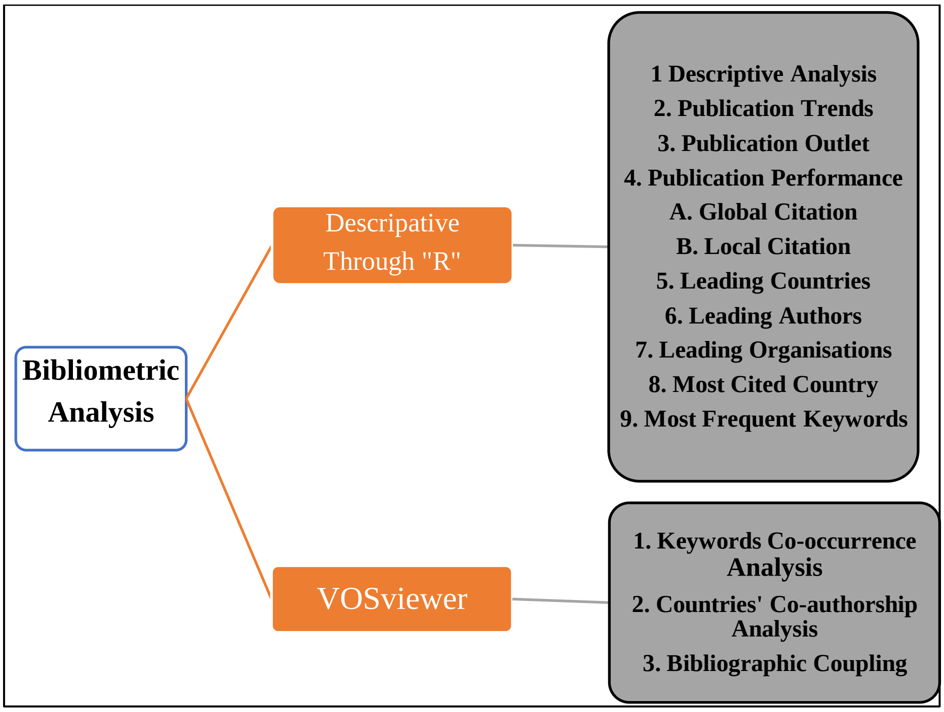


Figure No 3: Strategy for Bibliometric Analysis



4 Results

4.1 Descriptive analysis Through "R"

In this bibliometric analysis conducted using biblioshiny through R software, 572 documents were collected from the Scopus database. These documents comprised 429 articles and 143 conference papers published between 2012 and 2024. A total of 1589 authors contributed to these studies, these articles were published in 365 venues over 13 years. On average, each document received 14.99 citations. Authors utilized an average of 1502/572= 2.62 keywords per document. Each document had an average of 2.78 authors, indicating collaborative efforts. The cooperation index, at 2.88 per document, highlights multiple authors' collective research activity in the stock market sentiment using machine learning.

Table 1: Descriptive analysis

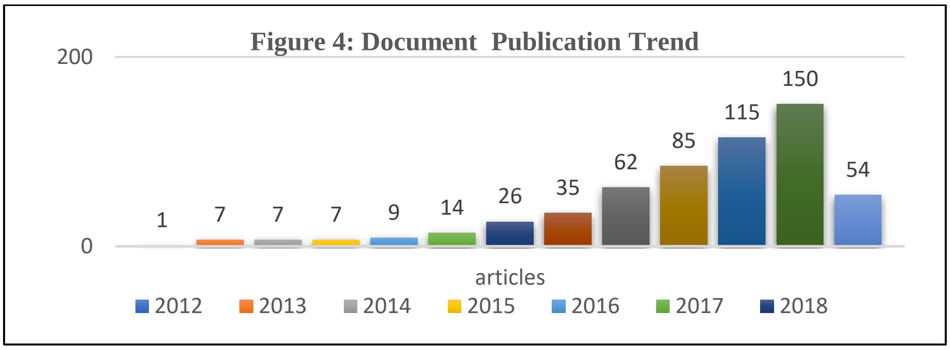
Main Information	**Description**	**Results**
Timespan	Year of publication	2012:2024
Sources (Journals, Books, etc)	Occurrence of documents in specific journals, books, etc.	365
Documents	Total number of studies	572
Total Citation		8575
Average citations per document	Total number of citations (26212) ÷ Total number of documents(1612)	14.99
Average citations per year per doc	---------------	2.81
References	Total references	28545
Document Types		
article	Total article	429
Conference paper	Total Conference paper article	143

Document Contents		
Keywords Plus (ID)	The total number of recurrent phrases identified automatically through the algorithm in the title of an article's references	2144
Author's Keywords (DE)	Total number of keywords used by authors	1502
Authors		
Authors	Total number of authors	1589
Author Appearances	Occurrence of authors in specific studies	1922
Authors of single-authored documents	Number of single authors per document	32
Authors of multi-authored documents	Number of authors of multi-authored documents	1557
Authors Collaboration		
Single-authored documents	Number of documents with a single author collaborated	32
Documents per Author	Total number of documents (572) ÷ Total number of authors(1589)	0.359
Authors per Document	Total number of authors (1589) ÷ Total number of documents(572)	2.78
Co-Authors per Documents	Author Appearances (1922) ÷ Total number of documents(572)	3.36
Collaboration Index	Total number of authors of multi-authored documents(1557) ÷ Total number of multi-authored documents (572-32) =540	2.88

Source: Data collected from Scopus using R bibliometric package

4.2 Publication trends

The distribution of articles by year of publication, as illustrated in **Figure 4**, indicates a significant increase in scholarly interest in sentiment analysis in the stock market over the past 13 years. The majority of research on this topic was published in 2023 (n = 150), with the highest increases observed in 2018 (85 percent) and 2020 (77 percent) compared to the previous year (n = 14 and 35, respectively). Surprisingly, before 2021, sentiment analysis in stock market research typically remained in the double and single digits only, but from 2022 onwards, it has consistently reached triple digits. The trend grew from 2016 onwards, and there was a notable increase. The growth trend peaked in 2022 and 2023.



Source: Calculation using "R" bibliometric package

4.3 Publication Outlet

Publication Outlet Journal Quality analysis (Top Journal)

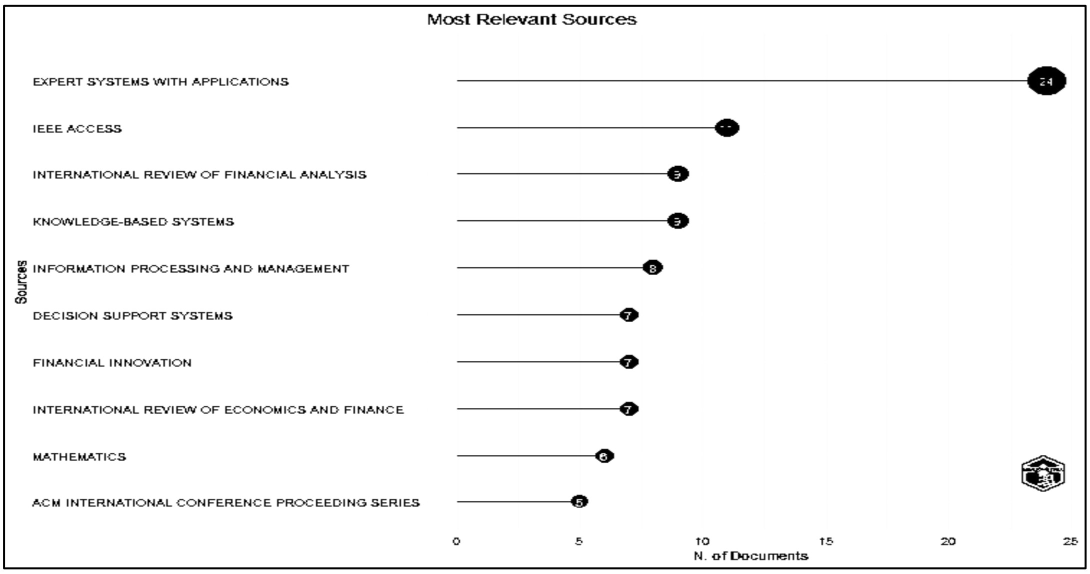
The distribution of articles by publication outlet shows that 'Expert System with Applications' is the most prolific source for stock market sentiment analysis using machine learning research (n=24 articles) **(Table no2)**. This is followed by IEEE access, (n=11 articles), International Review of Financial Analysis, Knowledge-Based Systems, Information Processing And Management**,** Decision Support Systems**,** and Financial Innovation, each hosting 11,9,9,8,7,7 articles on this area. Most publication outlets on this list are the journals that are ranked "A" by the Australian Business Deans Council. It is an assuring indicator for the future researches. The h-index, m-index, and g-index have been shown as given by bibliometric analysis in R software.

Table 2: Top 10 Contributing Journals

SOURCES	NP	TC	AC/PY	ABDC RANK	H INDEX	G INDEX	M INDEX	PUBLISHER	P/Y START
Expert Systems With Applications	24	715	29.79	C	11	24	1.1	Elsevier	2015
Ieee Access	11	47	4.27	A*	3	6	0.6	IEEE	2020
International Review Of Financial Analysis	9	50	5.55	A	4	7	1	Elsevier	2021
Knowledge-Based Systems	9	196	21.77	A	7	9	1.166	Elsevier	2019
Information Processing And Management	8	214	26.75	B	6	8	1	Elsevier	2019
Decision Support Systems	7	463	66.14	A*	7	7	0.7	Elsevier	2015
Financial Innovation	7	18	2.57	B	3	4	0.75	Business Perspectives Ltd	2021
International Review Of Economics And Finance	7	23	3.28	A	3	4	0.75	Elsevier	2021
Mathematics	6	23	3.83	C	3	4	0.75	Springer International Publishing	2021
ACM International Conference Proceeding Series	5	14	2.8		2	3	0.22	AssociationFor Computing Machinery	2016

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Figure: 5 Most relevant source



Source: biblioshiny R package

4.4 Publication performance

4.4.1 Global Citations

Global citations refer to the number of citations received without any filtration **(Kravet T., Muslu V).** In this review, the article with the highest number of global citations is '**Textual Risk Disclosures and Investors' Risk Perceptions**' (N= 304 citations), followed by '**Argument mining: a survey**' (*n* = 267 citations) (**see Table 3**). Remaining articles received less than 500 global citations.

Table: 3 Most cited articles are based on global citation documents rank-wise.

S. n.	Title	Author	Year	Total citations	Tc per year	Normalized tc
1	Textual Risk Disclosures And Investors' Risk Perceptions	Kravet T., Muslu V.	2013	304	25.33	3.75
2	Argument Mining: A Survey	Lawrence J.; Reed C.	2019	267	44.5	7.17
3	Sentiment Analysis Of Twitter Data For Predicting Stock Market Movements	Pagolu V.S.; Reddy K.N.; Panda G.; Majhi B.	2017	264	33	3.72
4	The Impact Of Microblogging Data For Stock Market Prediction: Using Twitter To Predict Returns, Volatility,	Oliveira N.; Cortez P.; Areal N.	2017	261	32.6	3.68

	Trading Volume, And Survey Sentiment Indices					
5	Temporal Relational Ranking For Stock Prediction	Feng F.; He X.; Wang X.; Luo C.; Liu Y.; Chua T.-S.	2019	245	40.83	6.58
6	Big Data: Deep Learning For Financial Sentiment Analysis	Sohangir S.; Wang D.; Pomeranets A.; Khoshgoftaar T.M.	2018	239	34.14	4.95
7	Natural Language Based Financial Forecasting: A Survey	Xing F.Z.; Cambria E.; Welsch R.E.	2018	239	34.14	4.95
8	Intraday Online Investor Sentiment And Return Patterns In The U.S. Stock Market	Renault T.	2017	203	25.37	2.86
9	Text Mining Of News-Headlines For Forex Market Prediction: A Multi-Layer Dimension Reduction Algorithm With Semantics And Sentiment	Khadjeh Nassirtoussi A.; Aghabozorgi S.; Ying Wah T.; Ngo D.C.L.	2015	195	19.5	5
10	Stream-Based Active Learning For Sentiment Analysis In The Financial Domain	Smailović J.; Grčar M.; Lavrač N.; Žnidaršič M.	2014	195	17.7	4.58

4.4.2 Local Citation

Local citations denote citations obtained from articles within the review corpus. Local citations are calculated based on references received from 572 articles focused on the " Bibliometric Analysis Of Stock Market Sentiments Using Machine Learning," sourced from Scopus and filtered for source quality, language, and subject relevance. Within this review, the article accruing the highest number of local citations is "Stock Market Sentiment Lexicon Acquisition Using Microblogging Data And Statistical Measures" (n = 63 citations) **(refer to Table 4),** followed by " The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume, and survey sentiment indices" (n = 35 citations). Subsequently, the articles " Intelligent Asset Allocation Via Market Sentiment Views," " Learning stock market sentiment lexicon and sentiment-oriented word vector from stock twits," BERT for stock market sentiment analysis" all are among the highest cited documents in stock market sentiment using machine learning.

Table: 4 Most cited articles based on local citations

S.N.	TITLE	AUTHORS	YEAR	LOCAL CITATIONS	GLOBAL CITATIONS	LC/GC RATIO (%)	NORMALIZED LC	NORALIZED GC
1	Stock market sentiment lexicon acquisition using microblogging data and statistical measures	Oliveira N.; Cortez P.; Areal N.	2016	63	127	49.60	7.76	5.34
2	The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume, and survey sentiment indices	Oliveira N.; Cortez P.; Areal N.	2017	35	261	13.40	5.69	3.68
3	Intelligent asset allocation via market sentiment views	Xing F.Z.; Cambria E.; Welsch R.E.	2018	25	90	27.77	7.38	1.86
4	Learning stock market sentiment lexicon and sentiment-oriented word vector from stock twits	Li Q.; Shah S.	2017	25	49	51.02	4.06	0.69
5	Bert for stock market sentiment analysis	Sousa M.G.; Sakiyama K.; Rodrigues L.D.S.;	2019	25	59	42.37	11.07	1.58
6	Natural language based financial forecasting: a survey	Xing F.Z.; Cambria E.; Welsch R.E.	2018	22	239	9.20	6.5	4.95
7	Intraday online investor sentiment and return patterns in the u.s. Stock market	Renault T.	2017	16	203	7.88	2.60	2.86
8	Big data: deep learning for financial sentiment analysis	Sohangir S.; Wang D.; Pomeranets A.; Khoshgoftaar T.M.	2018	14	239	5.85	4.13	4.95
9	Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation	Koratamaddi P.; Wadhwani K.;	2021	14	30	46.66	20.16	2.41

		Gupta M.; Sanjeevi S.G.						
10	Sentiment-aware volatility forecasting	Xing F.Z.; Cambria E.; Zhang Y.	2019	12	55	21.81	5.31	1.47

4.5 Leading Authors

Cambria E., authoreds12 articles, which include a total citation of 779, and (H index = 9) (G index =12) (M index =1.28) **(see table no 5**) followed by **Xing Fz**, with 5 articles having 494 total citations, **Chen Y** with 8 articles 191 citations, and **Liu Y**. with 6 articles and 394 total citations.

Table : 5 Top contributing authors

S.no.	Author	Np	Tc	H_index	G_index	M_index
1	Cambria E	12	779	9	12	1.28
2	Cortez P	5	416	5	5	0.55
3	Xing Fz	5	494	5	5	0.71
4	Chen Y	8	191	4	8	0.57
5	Eachempati P	4	64	4	4	1
6	Liu Y	6	394	4	6	0.66
7	Srivastava Pr	4	64	4	4	1
8	Tu W	5	71	4	5	0.44
9	Zhang Y	4	71	4	4	0.66
10	Chen C-C	4	43	3	4	0.6

4.6 Leading Organizations

The distribution of articles by institutions indicates that the University of Chinese Academy of Sciences, China, is the leading institution contributing to stock market sentiment using machine learning approach (*n* = 22 articles**) (Table 6**). This is followed by the Southwestern University Of Finance And Economics, china, Shanghai University of Finance and Economics, China, Nanyang Technological University, Singapore, and the University of Kwazulu-Natal, South Africa, each contributing 19, 18,17,12 articles respectively on the stock market sentiment using machine learning.

Table: 6 Top institutions in research

S. N.	AFFILIATION	COUNTRY	ARTICLES
1	University Of Chinese Academy Of Sciences	China	22
2	Southwestern University Of Finance And Economics	China	19
3	Shanghai University Of Finance And Economics	China	18
4	Nanyang Technological University	Singapore	17
5	University Of Kwazulu-Natal	South Africa	12

6	University Of Macau	Macau	12
7	Shandong Technology And Business University	China	11
8	Tsinghua University	China	11
9	Beijing	Beijing	9
10	Beijing Jiaotong University	Beijing	9

4.7 Leading Countries

The distribution of articles by countries indicates that authors from various countries have contributed to and published research on stock market sentiment using machine learning. The country-wise analysis reveals that China is the biggest contributor in this research with a frequency of 502. This is followed by India (280), USA (172), Brazil (43), Indonesia (42) and Iran (42) (**Figure 7**)

Figure: 7 Leading Countries

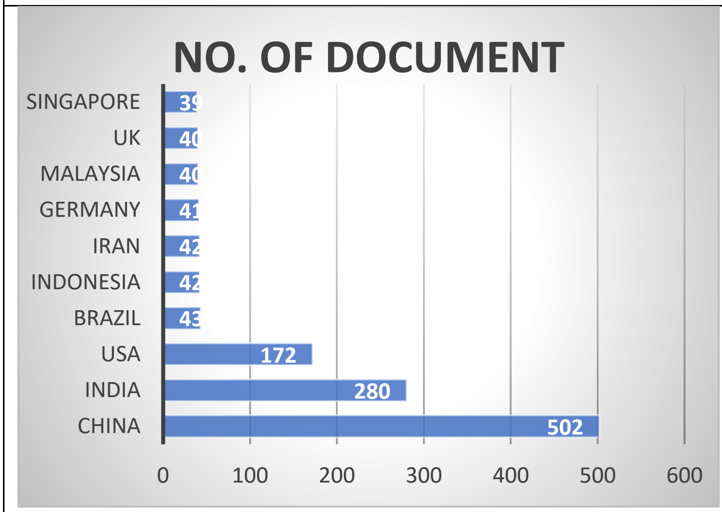


Table: 7 Most cited country in research

S. N.	COUNTRY	TC	ARTICLES	AVERAGE ARTICLE CITATIONS
1	China	1472	123	12
2	USA	924	32	28.9
3	Singapore	760	9	84.4
4	India	498	47	10.4
5	Portugal	436	6	72.7
6	United Kingdom	398	13	30.6
7	Australia	305	12	25.4
8	Malaysia	297	9	33
9	Ghana	287	2	143.5
10	Iran	273	13	21

4.9 Most Frequent Keywords based on Author keywords

The subsequent section analyzes the prevalence of keywords as indicated by authors. These keywords serve as optimal descriptors, encapsulating the essence of their respective works. They facilitate an enhanced comprehension of prevailing research trajectories and areas of interest within a specific domain of inquiry. **(Figure 4)** presents a visual depiction of these keywords through a word cloud, wherein larger font sizes correspond to greater frequency of occurrence. Among the keywords, "Sentiment analysis" emerges as the most recurrent term employed by authors. Additionally, terms such as "financial markets," "commerce," "investment," and "deep learning," all pertinent to the research on stock market sentiment using machine learning, feature prominently with high frequencies.

Figure: 8 Most frequent keywords based on author keywords

Table: 8 Top occurrences: Keywords

Srl no..	Keywords	Occurrences	Srl no..	keywords	Occurrences
1	Sentiment analysis	244	11	Data mining	53
2	Commerce	165	12	Long short-term memory	44
3	Financial market	159	13	Finance	41
4	Forecasting	125	14	Social media	41
5	Investment	125	15	Machine-learning	39
6	Deep learning	82	16	Learning algorithms	38
7	Social networking (online)	75	17	Machine learning	31
8	Learning system	65	18	Support vector machines	30
9	Electronic trading	60	19	Classification	29
10	Coasts	53	20	Natural language processing	26

5. **VOSviewer analysis:** It is performed on three bases - one is based on co-occurrence analysis**,** and another is co-authorship analysis and last one is on Bibliographic Coupling based on sources.

5.1 Co-occurrence analysis based on author keywords.

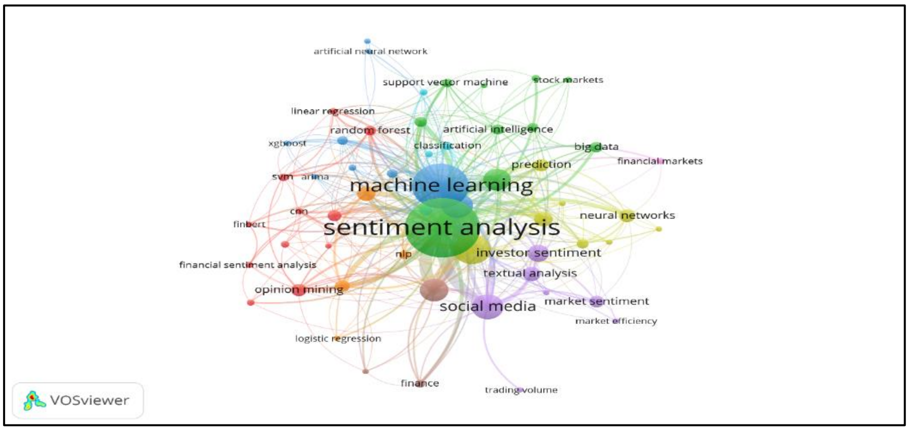
Steps of VOSviewer:

i) Type of Analysis: Co-occurrence, ii) Unit of Analysis: Author Keywords

Minimum number of occurrences of a keyword is 5 out of the 1504 keywords, of which 57 meet the threshold, for each of the 57 keywords. The total strength of the co-occurrence links with other keywords was calculated. The keyword with the greatest total link strength was selected and thus the number of keywords selected was found to be 57. The largest set of connected items consists of 57 items so we proceed to adopt 57 items only and maintain clusters on that basis. It can be divided into 9 clusters based on similar characteristics as mentioned below:

Cluster 1 (11 items), Cluster 2 (10 items), Cluster 3 (9 items), **Cluster 4 (8 items), Cluster 5 (7 items), Cluster 6 (4 items), Cluster 7 (4 item items), cluster 8 (3 items), cluster 9 (1 items)**

Figure 5: Maximum Occurance of Author Keywords



Source: VOSviewer package

Cluster 1(11 items)	**Cluster 2(10 items)**	**Cluster 3 (9 items)**	**Cluster 4(8 items)**	**Cluster 5 (7 items)**
CNN	Artificial intelligence	ARIMA	Deep learning	Behavioural finance
Financial sentiment analysis	Big data	Artificial neural network	Forecasting	Investor sentiment
Finbert	Deep reinforcement learning	Data mining	Neural network	Market efficiency
Linear regression	Financial market	Machine learning	Neural networks	Market sentiment
Opinion mining	Portfolio selection	Portfolio optimization	Prediction	Social media
Random forest	Sentiment analysis	Stock market	Stock price	Textual analysis
Sentiment classification	Stock market	Stock price prediction	Technical analysis	Trading volume
Sentiment lexicon	Stock prediction	Time series	Twitter	Cluster 6 (4 items)
Stock market prediction	Support vector machine	xgboost	Cluster 8 (3 items)	Classification
SVM	Text mining	Cluster 7 (4 items)	Finance	Financial news
Time series	Cluster 9 (1 item)	BERT LSTM	Natural language	Optimization

analysis				processing	
	Financial market	NLP	Logistic regression	Text classification	Sentiment

5.2 Co-authorship analysis based on Countries:

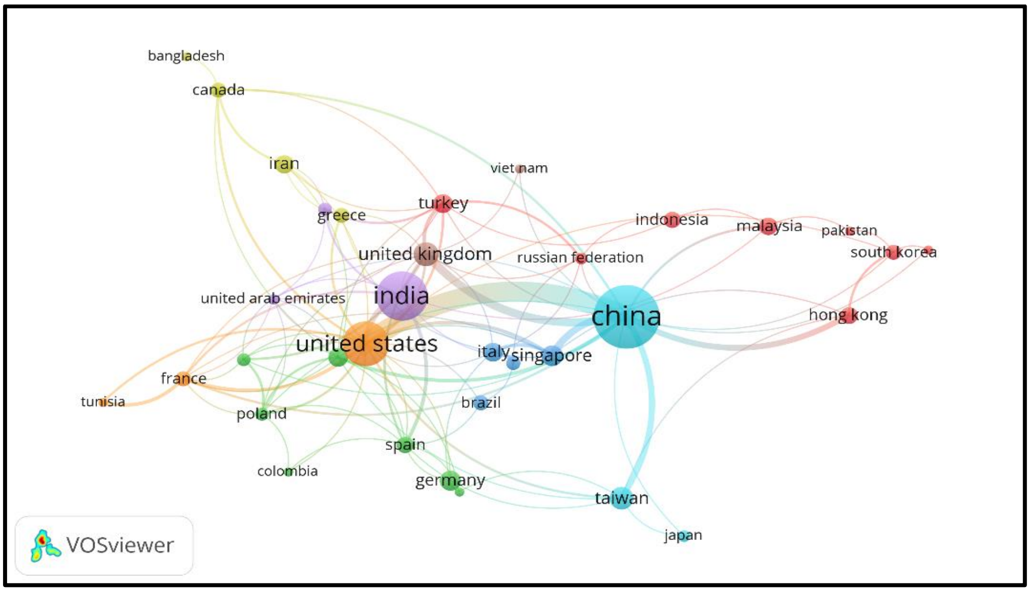
Step of VOSviewer:

i)Type of Analysis: Co-authorship, ii) Unit of Analysis: Countries

Minimum number of documents of a country is 5 and Minimum number of citations of a country is 0. 34 countries out of 78 countries meet the thresholds. For each of the 34 countries, the total strength of the co-authorship links with other countries was calculated. The largest set of connected items consists of 34 items, so we adopt 34 items only and maintain clusters on that basis. It can be divided into 8 clusters based on similar characteristics as mentioned below:

Cluster 1 (8 items), Cluster 2 (7 items), Cluster 3 (4 items), Cluster 4 (4 items), Cluster 5 (3 items), Cluster 6 (3 items), Cluster 7 (3 items), cluster 8 (2 items).

Figure 7: Co-authorship analysis based on Countries:



Source: VOSviewer package

Clusters Co-authorship analysis based on Countries:

Cluster 1(8 Items)	**Cluster 2(7 Items)**	**Cluster 3 (4 Items)**	**Cluster 4(4 Items)**
Hong Kong	Australia	Brazil	Bangladesh
Indonesia	Colombia	Italy	Canada
Malaysia	Germany	Portugal	Greece
Pakistan	Netherland	Singapore	Iran
Russian Federation	Poland	**Cluster 6 (3 Items)**	**Cluster 5 (3 Items)**
South Korea	South Africa	China	India
Thailand	Spain	Japan	Saudi Arabia
Turkey	**Cluster 7 (3 Items)**	Taiwan	United Arab

					Emirat
Cluster 8 (2 Items)		**United States**			
United Kingdom	Vietnam	France	Tunisia		

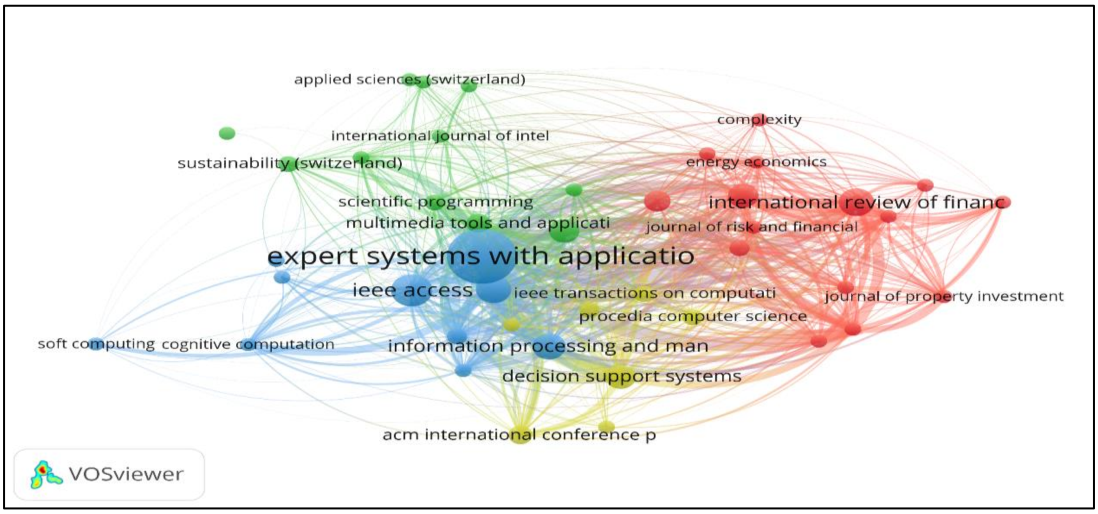
5.3 Bibliographic Coupling based on Sources (Journal)

Step of VOSviewer:

i) Type of Analysis: Bibliographic Coupling, ii) Unit of Analysis: Sources: The minimum number of documents of a source is 3 Minimum number of citations of a source is 0. 43 source out of the 365 meet the thresholds. For each of the 43 sources, the total strength of the bibliometric coupling links with other sources was calculated. The sources with the greatest total link strength was selected and thus the number of sources selected was found to be 43. The largest set of connected items consists of 43 items, so we adopt 43 items only and maintain clusters on that basis. It can be divided into 4 clusters based on similar characteristics as mentioned below:

Cluster 1 (15 items), Cluster 2 (11 items), Cluster 3 (10 items), Cluster 4 (7 items).

Figure 8: Bibliographic Coupling based on Sources (Journal)



Source: VOSviewer package

Clusters Bibliographic Coupling based on Sources (Journal)

Cluster 1 (15 Items)	**Cluster 2 (11 Items)**
Accounting And Finance	Applied Sciences (Switzerland)
Complexity	Computers, Materials And Continua
Energy Economics	Engineering Letters
Finance Research Letters	Financial Innovation
Frontiers In Artificial Intelligence	Intelligent Systems In Accounting, Finance And Management
International Review Of Economics And Finance	International Journal Of Intelligent Systems And Applications In Engineering
International Review Of Financial Analysis	Multimedia Tools And Applications
Journal Of Business Research	Neural Computing And Applications

Journal Of Corporate Finance	Scientific Programming
Journal Of International Financial Markets, Institutions And Money	Sustainability (Switzerland)
Journal Of Property Investment And Finance	Wseas Transactions On Systems
Journal Of Risk And Financial Management	**Cluster 3 (10 Items)**
Mathematics	Artificial Intelligence Review
North American Journal Of Economics And Finance	Ceur Workshop Proceedings
Research In International Business And Finance	Cognitive Computation
Cluster 4 (7 Items)	Expert Systems With Applications
Acm International Conference Proceeding Series	Ieee Access
Decision Support Systems	Industrial Management And Data Systems
Ieee International Conference On Industrial Informatics (Indin)	Information Processing And Management
Ieee Transactions On Computational Social Systems	Journal Of Intelligent And Fuzzy Systems
Peerj Computer Science	Knowledge-Based Systems
Procedia Computer Science	Soft Computing
Proceedings Of Spie - The International Society For Optical Engineering	

Source: Author elaboration using VOSviewer package

Conclusion

While the literature on stock market sentiment analysis using machine learning is extensive, it has been noticed that a gap in the bibliometric study of stock market sentiment analysis using machine learning exists despite the abundance of literature on this topic. To address this gap, our study examined 572 research publications from 2012 to April 2024. Our analysis of this data identifies the most impactful authors, such as ‘**Cambria E**’ and a prominent journal is **‘Expert Systems with Applications’**. We observe a rising publication trend, with the highest number of articles published in **2022** and **2023**. The analysis also reveals global and local citation performance, with the University of **‘Chinese Academy of Science** leading among top institutions. **‘India’** is the second most cited country after **‘China’** and in this study we analyse **‘Sentiment analysis’** as the most frequent keyword. We conducted network analyses using three approaches: co-occurrence analysis based on author keywords (resulting in 9 clusters of different characteristics), co-authorship analysis based on countries (with 8 clusters), and bibliographic coupling based on sources (resulting in 4 clusters).

There are certain limitations of this study. First, Scopus is our only primary database for collecting research articles. Further studies can focus on and extend the study utilizing other bibliometric and citation databases like WOS and Google Scholar etc. and further studies can also be conducted by combining two databases like Scopus and WOS. Secondly, this study focused on a limited timeframe 2012-2024. Third, our study is limited to subject areas of economics, econometrics, finance, business

management, accounting, computer science, and engineering, while some other areas also focus on machine learning techniques. So future studies can focus on other subject areas also like decision making etc. For network analysis, although we used co-occurrence, bibliographic coupling, and co-authorship analysis, some other techniques can also be used in future studies.

References

1. hardwaja, A. (2015). Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty. *Procedia Computer Science*.
2. Bodla1, S. a. (2023). Sentiment Analysis Using Machine in Stock Market: A Bibliometric Visualization. *sage*.
3. Duyu Tang, B. Q. (2015). Deep learning for sentiment analysis: successful approaches and future challenges. *Data Mining and Knowledge Discovery*.
4. FELDMAN, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*.
5. Gao, Y. (2010). Sentiment classification for stock news. *ICPCA10 - 5th International Conference on Pervasive Computing and Applications*.
6. Henrique, B. M. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*.
7. Inani, S. K. (2024). Navigating The Technical Analysis In Stock Markets: Insights From Bibliometric And Topic Modeling Approaches. *Investment Management and Financial Innovations*.
8. Lee, B. P. (2002). Thumbs up? Sentiment Classification using Machine Learning. *(EMNLP)*.
9. Li Im, T. (2013). Analysing market sentiment in financial news using lexical approach. *2013 IEEE Conference on Open Systems, .*
10. Liu, B. (2012). *Sentiment Analysis and opinion mining.* morgan & claypool.
11. Naveen Donthu, s. k. (2021). How to conduct a bibliometric analysis: An overview and guidelines . *Journal of Business Research*.
12. Pagolu, V. S. (2016). Sentiment analysis of Twitter data for predicting stock market movements. *International Conference on Signal Processing, Communication, Power and Embedded System, .*
13. Pooja Bagane, N. M. (2021). Bibliometric Survey for Stock Market Prediction using Sentimental analysisand LSTM . *Library Philosophy and Practice (e-journal)*.
14. Porshnev, A. (2013). Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis. *Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013*.
15. Zulfikar, Z. (2022). “Bibliometric analysis of stock market performance throughout the COVID-19. *"Investment Management and Financial Innovations"*.
16. Das, S., Chowdhury, U., Lijin, N. S., Deep, A., Saha, S., & Maurya, A. (2024). Investigate How Market Behaves: Toward an Explanatory Multitasking Based Analytical Model for Financial Investments. *IEEE Access*, *12*, 30928-30940.
17. Chong, K., & Shah, N. (2022). Comparison of naive bayes and SVM classification in grid-search hyperparameter tuned and non-hyperparameter tuned healthcare stock market sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *13*(12).
18. Patil, D., Patil, S., Patil, S., & Arora, S. (2022). Financial Forecasting of Stock Market Using Sentiment Analysis and Data Analytics. In *Intelligent Sustainable Systems: Selected Papers of WorldS4 2021*,

Volume 2 (pp. 423-430). Springer Singapore.
19. Gurav, U., & Kotrappa, D. (2020). Predict stock market's fluctuating behaviour: role of investor's sentiments on stock market performance. *SSRG International Journal of Engineering Trends and Technology*, *68*, 72-80.
20. Gurav, U. P., & Kotrappa, S. (2020). Sentiment aware stock price forecasting using an SA-RNN-LBL learning model. *Engineering, Technology & Applied Science Research*, *10*(5), 6356-6361.
21. Alazba, A., Alturayeif, N. S., Alturaief, N., & Alhathloul, Z. (2020). Saudi Stock Market Sentiment Analysis using Twitter Data. In *KDIR* (pp. 36-47).
22. Sangsavate, S., Tanthanongsakkun, S., & Sinthupinyo, S. (2019, November). Stock market sentiment classification from FinTech News. In *2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE)* (pp. 1-4). IEEE.
23. Rajput, V. S., & Dubey, S. M. (2016, October). Stock market sentiment analysis based on machine learning. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)* (pp. 506-510). IEEE.
24. Lin, N., Xu, W., Zhang, X., & Lv, S. (2014). Can web news media sentiments improve stock trading signal prediction?.
25. Farman, A. L. I., & Pradeep, S. U. R. I. (2022). A Bibliometric Analysis of Artificial Intelligence-Based Stock Market Prediction. *The Eurasia Proceedings of Educational and Social Sciences*, *27*, 17-35.
26. Vicari, M., & Gaspari, M. (2021). Analysis of news sentiments using natural language processing and deep learning. *AI & society*, *36*(3), 931-937.
27. Long, W., Gao, J., Bai, K., & Lu, Z. (2024). A hybrid model for stock price prediction based on multi-view heterogeneous data. *Financial Innovation*, *10*(1), 48.
28. Liu, Q., Son, H., & Lee, W. S. (2024). The game of lies by stock investors in social media: a study based on city lockdowns in China. *Financial Innovation*, *10*(1), 65.
29. Ma, Y., Wu, P., Ling, C., & Ding, S. (2024). Research on public opinion effecting on stock price during crises based on model checking. *Expert Systems with Applications*, 123442.
30. Zhang, Z., Goodell, J. W., Shen, D., & Lahmar, O. (2024). Media opinion divergence and stock returns: Evidence from China. *International Review of Financial Analysis*, 103140