Minnesota Journal of Business Law and Entrepreneurship

Volume 2026, No. 2

ISSN: 1540-3270

Exploring Global Scholarly Contributions to AI-Driven Stock Prediction through Analytics: A Bibliometric Evaluation

[1]Dr. Christo Aditya Bikram Bepari, [2]Dr. Suresh Chandra Ch, [3]Dr. Manoranjan Dash, [4]Dr. Bibhuti Bhusan Pradhan

[1]Assistant Professor, Department of Commerce and Management Studies, Vignan's Foundation for Science Technology and Research (Deemed to be University), Hyderabad, Telangana, India.

[2]Associate Professor, Department of Marketing, Vignana Jyothi Institute of Management, Pragathi Nagar, Hyderabad Telangana, India.

[3]Professor, Faculty of Management Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India

[4]Professor, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India.

Abstract

The financial market plays a pivotal role in global economy and is a domineering field of research. It not only provides corporates a way to raise capital to fund their operations and expansion plans but additionally, provides investors with an opportunity to earn a return on their investments and grow their wealth over time. Although the massive usage of machine learning techniques in stock market field is up trending, there has been little focus on developing a framework that synthesizes the primary trends and research on the topic. This study is one of the attempts to thoroughly investigate the published literatures in few decades and focuses on addressing this gap by conducting further analysis of complementary bibliographic sources, evaluating 834 reputable bibliometric studies published in the Scopus database. A total of 2,037 authors contributed to publications, with 58 authors writing 68 documentsindividually and 19.78% of these authors have engaged in international collaborations. The average number of citations per document is 24.02. The leading nations with greatest number of publicationsare CHINA, INDIA and US ,i.e., with 238,210 and 82 articles respectively. The prominent journal is 'Expert Systems with Applications' (with 57 articles) and the most productive authors areLI X and WANG X (with 14 articles each) inthe Scopus database. This research reveals that as subset of AI (Artificial Intelligence), the ML (Machine Learning) and the DL(Deep Learning) application in stock market is evolving and expanding. This paper enhances the existing literature by offering a thorough overview of the articles that focuses on application of AI and ML in stock market prediction.

Keywords: Machine Learning, Bibliometrics, Stock market, Artificial Intelligence, Science mapping, knowledge management

1 Introduction

The use of machine learning methods is becoming increasingly common across various disciplines of the stock market. To date, several machine learning methodsare in implementation to tackle stock market prediction tasks, including forecasting stock prices, identifying market trends, detecting anomalies, and optimizing portfolio management. Numbers of studies on emerging AI-based approaches have have shown remarkable potential in enhancingthe reliability and fitness of stock market prediction models (Ghosh et al.,2024; Rath et al.,2024;Dang, Q. V. ,2019; Zhang & Lei,2022).

The stock market is a platform where the trading of shares and securities takes place regularly. Due to its inherent volatility, experts often consider it a high-risk investment.The stock market system is notoriously intricate due to the perpetual interplay of dynamic trends, volatility, and the inherent noise generated by numerous market participants, news, and events, making it a challenging and unpredictable landscape for investors (Abdullah et al., 2024). Accurately predicting the stock market is ainspiriting yet puzzling endeavor in the ever-changing financial landscape. Earlier, analysts were dependent on conventional approaches such as "*AR (Auto regressive),ARMA (Auto Regressive Moving Average),ARCH (Autoregressive Conditional Heteroscedasticity),ARIMA (Auto Regressive Integrated Moving Average),GARCH (Generalized Auto Regressive Conditional Heteroscedasticity) and Stochastic Volatility (SV) etc.",*to forecast prices using

historical data(Cheng et al.,2023; Lin, Z.,2018;Gaire,H.N., 2017; Saini& Mittal,2014). Nevertheless, the effectiveness of these strategies have been diminished (Vuong et al.,2024; Nassar et al.,2020; Sun et al.,2022;Zhong &Enke, 2019; Bansal et al.,2022)due to the complexity of the financial market. Recently, many researchers have undertaken research studies on the sets of machine learning models encompassing linear models, traditional machine learning, deep learning, reinforcement learning, and an integration of all models, and how these methods can be adapted to accurate predictions with a low probability of errors (Mukherjee et al.,2023).

Ayyildiz&Iskenderoglu (2024)applied machine learning algorithms to predict the fluctuations of major stock market indices, specifically focusing on (G-7 country)"NYSE 100, NIKKEI 225, FTSE 100, CAC 40, DAX 30, FTSE MIB, and TSX". They noted that artificial neural network gave the best forecast results for "NYSE 100, DAX 30, FTSE 100, and FTSE MIB", while logistic regression proved more accurate for "CAC 40, NIKKEI 225, and TSX".Sheth& Shah (2023)recently used the strength of several machine learning and deep learning models such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Long Short-Term Model (LSTM) to assess historical stock prices and to develop a predictive model. However, a related study conducted by Hiransha et al. (2018)were employed four DL-techniques "LSTM, MLP(Multi-Layer Perceptron), CNN(Convolutional NN), and RNN(Recurrent NN)" to forecast the stock price of "MARUTI, AXIS BANK, HCL(from NSE, India)" and the stock price of "CHESAPEAK ENERGY, BANK OF AMERICA(from NYSE)" based on historical prices.Sivadasan et al. (2024) suggested to the usage of GRU (Gated Recurrent Unit) and LSTM models.requires the use of technical consideration for all of the models be utilized to apply all of the models and to provide the models with a direction on when it is appropriate to outperform the models.Zou et al. (2024) has developed a cutting-edge stock trading method leveraging Cascaded LSTM with Proximal Policy Optimization (CLSTM-PPO) system which effectively captures concealed daily Stock Market Data and surpasses existing models in various parameters, including cumulative returns, max.earning rate,and mean profitability per transaction.

Although there is significant interest and adoption of machine learning and artificial intelligence in the finance industry (Jain et al.;2024), there has been limited focus for integrating and synthesizing the disparate research streams and findings into a comprehensive framework.Scholars and researchers believe that emergence of AI and ML in FinTechwill play crucial part towards building more efficient and effective strategy for financial market(Jrad, M.,2023; Arsenyan& Piepenbrink,2023).In order to address the existing knowledge gap, this research aims to supplement existing studies by exploring complementary sources of bibliographic information, thus providing a more complete understanding of the topic and filling the gaps in current literature. Bibliometrics is is widely considered for usage in various disciplines as it offersvaluable perspectives on the current landscape of research, shedding light on influential studies, emerging trends, and potential knowledge gaps that warrant further exploration.

The primary aim of this research is to perform a comprehensive bibliometric analysis to investigate the research landscape concerning the application of machine learning in stock market prediction. We employed a mixed approach that combines performance analysis and conceptual mapping to identify key trends, patterns, relationships within the literature and answer four key questions: (i) What is the annual growth rate of publications on employing AI and MLmethods in stock market prediction field? (ii) Which journals, documents, authors, and countries have had the most significant impact on employing AI and MLmethods in stock market prediction field? (iii) What are the current gaps and promising research directions in the application of artificial intelligence and machine learning techniques to stock market prediction? (iv)How have research topics and trends evolved over time on employing AI and MLmethods in stock market prediction field?

s2 Methodology

Bibliometric analysis is a scientific field that employs statistical techniques to quantify and visualize the evolution of knowledge in a particular field. Its use has been growing within the scientific community as a valuable tool (Ellegaard,O.,2018).Itserves to measure quality and impact of various works within that domain.It equips scholars to attain a comprehensive overview for gaining a deep understanding of the topic, identify gaps in knowledge, and generate new research ideas (Lim & Kumar,2024;Donthu et al.,2021).

The data for this study were sourced from the citation database Scopus. Scopus is a comprehensive bibliographic database with abstracts and citations from journal articles, conference proceedings, conference

papers, and books across worldwide and a wide range of disciplines. Scopus has great search facilities, including the use of Boolean operators, wildcards, and proximity operators to refine the search criteria, and allow users to identify relevant publications with precision.In order to identify relevant articles, a comprehensive search was conducted in the Scopus database using keywords specifically tailored to machine learning applications and stock market industry.

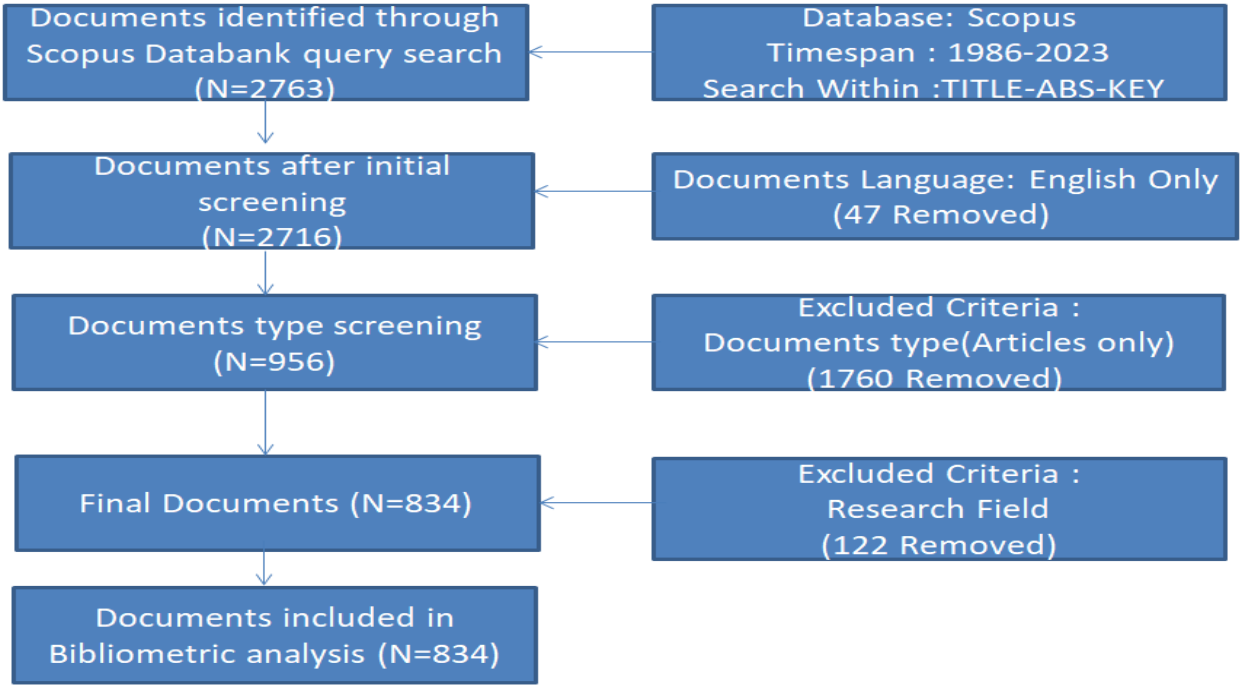

Figure-1: Document Selection Procedure (PRISMA Chart)

We stringently followed the guidelines of "PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analyses)", to ensure consistency and transparency in the identification and synthesis of studies(Rethlefsen et al., 2021), thereby ensuring the reliability and validity of the outputs.To point out relevant works on intersection of financial markets and machine learning, we used the following search string in Scopus: [(TITLE-ABS-KEY "machine learning" OR "deep learning" OR "artificial intelligence" OR "data science" OR "algorithm" OR "Neural network" OR "data analytics" OR "automation" OR "robot" OR "big data" OR "text mining" OR "natural language processing" OR "soft computing") AND (TITLE-ABS-KEY "stock market prediction" OR "stock* predict*" OR "stock market forecast" OR "stock* forecast*" OR "stock market trend" OR "stock market analysis" OR "stock volatile*" OR "stock* manipulate*")].This search initially yielded a collection of 2,763 documents. Subsequently, filters were applied to refine the results, focusing on language (English), which narrowed the selection down to 2,716 articles and then document type (articles only) which further narrowed the selection down to 956. These remaining articles were then meticulously reviewed, with their titles and abstracts scrutinized to identify those most relevant to our study and finally 834 articles were selected for input. The data provided by Scopus includes bibliographic information such as author names, affiliations, publication titles, abstracts, publication dates, citation counts, and more. The tabular data in the form of .csv file downloaded. The data was converted to valuable visual information along with variables needed to interpret the results. We leveraged the capabilities of R-Bibliometrix and VOS viewer software, which provide robust tools for statistical analysis and visualization, to execute a range of statistical analyses and generate high-quality graphical outputs, aiding in the exploration and interpretation of our research findings.

3 Results

We may broadly categories the bibliometric analysis into two analytical components. One is the (i) performance evaluation and other (ii) conceptual framework. While the first examines the contributions of various research components, the second deals with network of co-occurrences. Through a rigorous analysis of performance metrics and scientific mapping techniques, we can pinpoint key indicators such as influential authors, documents, countries and high impact journalsetc. may be identified. On the other hand, conceptual structure analysis helps to focus on the thematic evolution and intellectual structure of the field, identifying Co-occurrence frequency of scientific domain-specific words.

3.1 Performance analysis

We conducted a descriptive statistical analysis for the first research question [Investigating the annual growth rate of publications on employing AI and ML methods in stock market prediction field]. The results are presented in Table 1, which reveals that over the past 37 years (1986-2023), a sum of 834 articles have been published in the Scopus, averaging 24.02 citations per document.

Table-1: Description Statistical analysis of Research Questions

Sl.No.	Description	Results
1	MAIN INFORMATION ABOUT DATA	
2	Timespan	1986:2023
3	Sources (Journals, Books, etc)	385
4	Documents	834
5	Annual Growth Rate %	7.96
6	Document Average Age	6.72
7	Average citations per doc	24.02
8	References	29520
9	DOCUMENT CONTENTS	
10	Keywords Plus (ID)	3196
11	Author's Keywords (DE)	1988
12	AUTHORS	
13	Authors	2037
14	Authors of single-authored docs	58
15	AUTHORS COLLABORATION	
16	Single-authored docs	68
17	Co-Authors per Doc	3.17
18	International co-authorships %	19.78
19	DOCUMENT TYPES	
20	article	834

Source: compiled from R-Bibliometrix and VOS viewer

The findings reveal that from 1986 to 2023, 2037 distinct authors have contributed to the studies dealing with applying machine learning for stock market prediction. Among them, 58 authors have individually produced 68 documents. Additionally, 19.78% of the authors have engaged in international collaborations. On average, each author contributed 0.388 documents, while each document was co-authored by an average of 3.17 researchers. These metrics underscore the significant advancement in machine learning and artificial intelligence studies for stock market prediction.

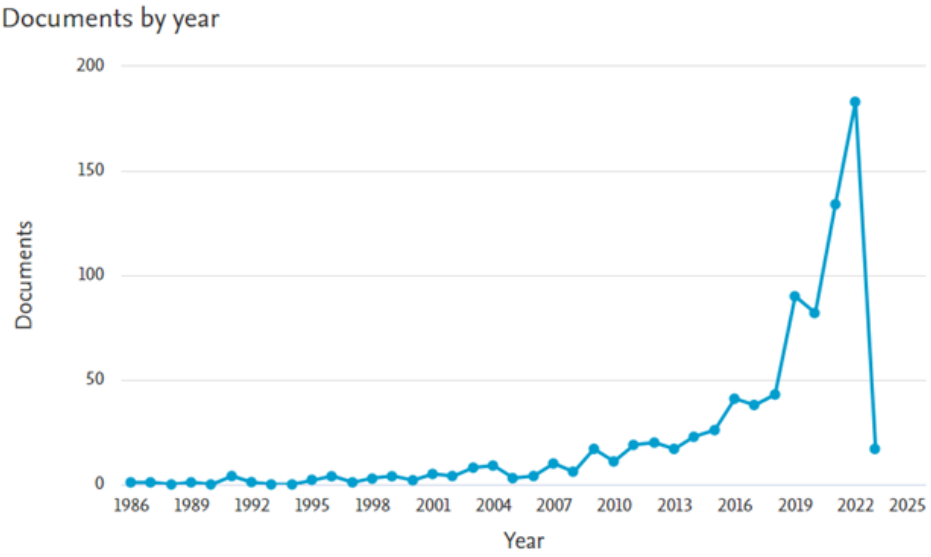

Figure-2: Year wise number of Scientific publications

Figure 2 shows the annual distribution of scientific publication. Analyzing the annual distribution of publications offer valuable insights into the growth and development of the topic, the trends and patterns over time. We can observe, the number of publications is increasing with only a few decline points in last decade which suggests that the field is in growing trend. The impact of sharp rise and fall in the pattern is general, because of Covid in 2019-20. To address the next research question that deals with investigation on journals, documents, authors, and countries having most significant impact on employing AI and MLmethods in stock market prediction field, we identified and document below the distribution of authors, documents, journals, and countries.

3.1.1. Performance of authors and Co-authorship network analysis:

Figure 3 illustrates the top 10 authors with the highest number of published articles. The data indicates that LI X and Wang X are the leading authors, each having published 14 articles. They are followed by Wang J, Zhang Y, and Wang H, with 13, 11, and 10 articles respectively. The output of other authors has been modest, with nine or fewer articles.

Figure-3: Top10 authors with the highest number of published articles

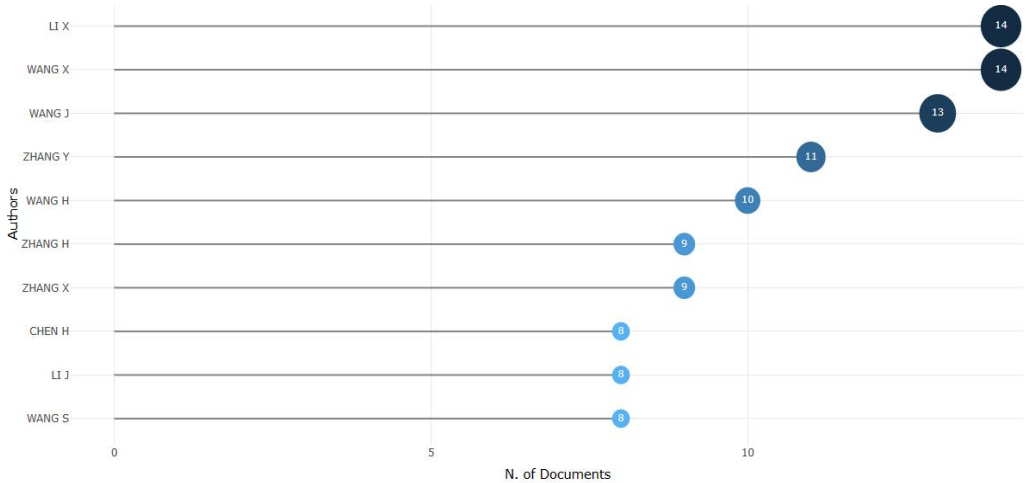

Table 3 Authors' publishing frequency (Lotka law)

Documents written	N. of Authors	Proportion of Authors
1	1713	0.841
2	204	0.100
3	58	0.028
4	30	0.015
5	15	0.007
6	3	0.001
7	3	0.001
8	4	0.002
9	2	0.001
10	1	0.000
11	1	0.000
13	1	0.000
14	2	0.001

Source: Compiled from R-Bibliometrix and VOS viewer

Table 3 indicates that a significant majority (84.1%) of authors have written one paperonly, that aligns with Lotka's Law. Authors publishing two papers is significantly lower compared to those publishing just one paper, i.e., approximately 12% of the number of single-paper authors.

Table -4: Productive Authors Analysis:

Author Name	Number of Articles	T.C	A.C	TLS
Li X.	14	290	20.71	7
Wang X.	14	240	17.14	6
Wang J.	13	146	11.23	9
Zhang Y.	11	152	13.82	7
Wang H.	10	201	20.10	8
Zhang H.	9	122	13.56	5
Zhang X.	9	99	11.00	5
Chen H.	8	859	107.38	3
Li J.	8	95	11.88	3
Wang S.	8	287	35.88	4
Chen Y.	7	195	27.86	2
Zhang W.	7	47	6.71	4
Liu X.	6	26	4.33	4
Wang L.	6	87	14.50	3
Wang Y.	6	79	13.17	3
Abraham A.	5	161	32.20	1
Chen W.	5	112	22.40	3
Li Y.	5	52	10.40	2

Wang W.	5	142	28.40	2
Wu S.	5	49	9.80	5
Yang Y.	5	95	19.00	2
Zhang J.	5	33	6.60	3
Zhang S.	5	144	28.80	4
Zhao C.	5	105	21.00	2
Zhao J.	5	313	62.60	1

Table 4 provides additional support to the study on account of published articles, T.C., A.C., TLS. Li X and Wang X are the top productive authors with 14 articles each and corresponding T.C., A.C., TLS of 290,20.71,7 for Li X and T.C., A.C., TLS of 240,17.14,6for Wang X.Figure 4 depicts the network showcasing the co-authorship connections among prominent researchers and helps in recognizing influential researchers, understanding the collaborative landscape, and fostering new research partnerships.

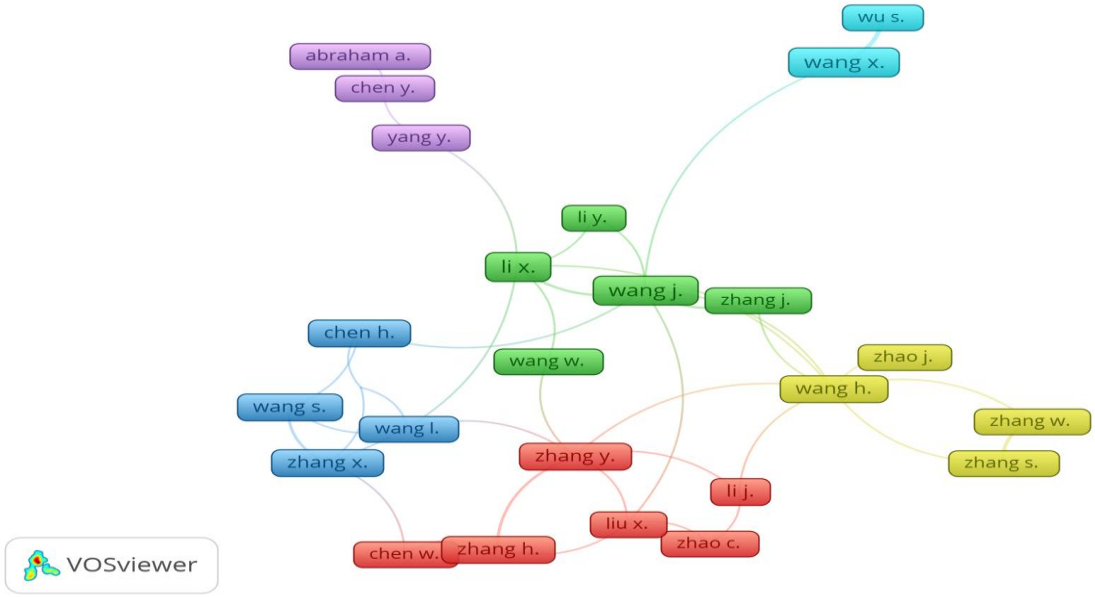

Figure-4: Collaboration network of authors

3.1.2Top research outlets

In academia, research outlets serve as platforms for scholars to disseminate their findings to the broader scientific community. These platforms disseminate state-of-the-art research, varying in focus, depth, and noteworthy criteria. Essential journals play a pivotal role in scrutinizing trends within specific fields of interest. This study employed VOS viewer to systematically chart significant journals, utilizing suitable indicators and metrics. To accomplish this, the metric parameters were established at a minimum of five articles and forty citations. They were not tied to specific benchmarks. Instead, they were determined through a synthesis of prior scientometric evaluations, supplemented by the authors' own judgment and expertise (Bello et al., 2021). The creation of a network map highlighting the leading journals was facilitated by these parameters. The summarized findings are presented at Table 5, including metrics [such as the count of articles, T.C., A.C., TLS]. When gauging productivity through publication frequency, "Expert Systems with Applications" and "IEEE Access" emerged as the top two research outlets, with 57 and 22 outputs, respectively.

Table -5 Most Relevant Journals

Name of the Journal	No of Articles	Total Citations	Avg. citations	Total link strength	Norm. citations
Expert systems with applications	57	3816	66.9474	88	131.38
Ieee access	22	409	18.5909	24	24.80
Neurocomputing	20	1909	95.45	56	36.41
International journal of recent technology and engineering	14	63	4.5	2	3.30
Soft computing	14	191	13.6429	15	11.58
Applied soft computing journal	13	793	61	38	22.83
Neural computing and applications	13	447	34.3846	30	24.84
Applied soft computing	10	94	9.4	24	20.84
Computational intelligence and neuroscience	9	61	6.7778	11	3.49
Information sciences	9	268	29.7778	7	17.64
International journal of intelligent engineering and systems	8	60	7.5	6	1.59
Mathematical problems in engineering	8	148	18.5	7	5.67
Knowledge-based systems	7	841	120.1429	25	21.04
Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)	7	134	19.1429	2	3.18
Indian journal of science and technology	6	45	7.5	2	1.41
Multimedia tools and applications	6	157	26.1667	12	8.61
Plos one	6	44	7.3333	5	2.59
Financial innovation	5	133	26.6	11	9.47
Journal of theoretical and applied information technology	5	54	10.8	3	1.46

Source: Compiled from R-Bibliometrix and VOS viewer

As depicted in Figure 5, generated using VOSviewer's source citation network, node side corresponds to citationsnumber, serving as an indicator of prominent outlets' impact within our selected field of study. Larger nodes signify greater impact. In Fig. 5, the node corresponding to "Expert Systems with Applications" stands out with a noticeably larger size, indicating a substantial impact in this domain, with a cumulative citation count of 3816. Outlets of a similar nature are identified by shared colors and lines represent connections strength through cross-citation. Let’s take an example, within green cluster, "Expert Systems with Applications" demonstrates notably higher citation connections with "Neurocomputing", "Lecture notes in computer science” and "Indian Journal of Science & Technology". This suggests frequent cross-referencing among these publications across various contexts. Furthermore, the clustering of similar outlets may stem from shared research focuses guiding their content.

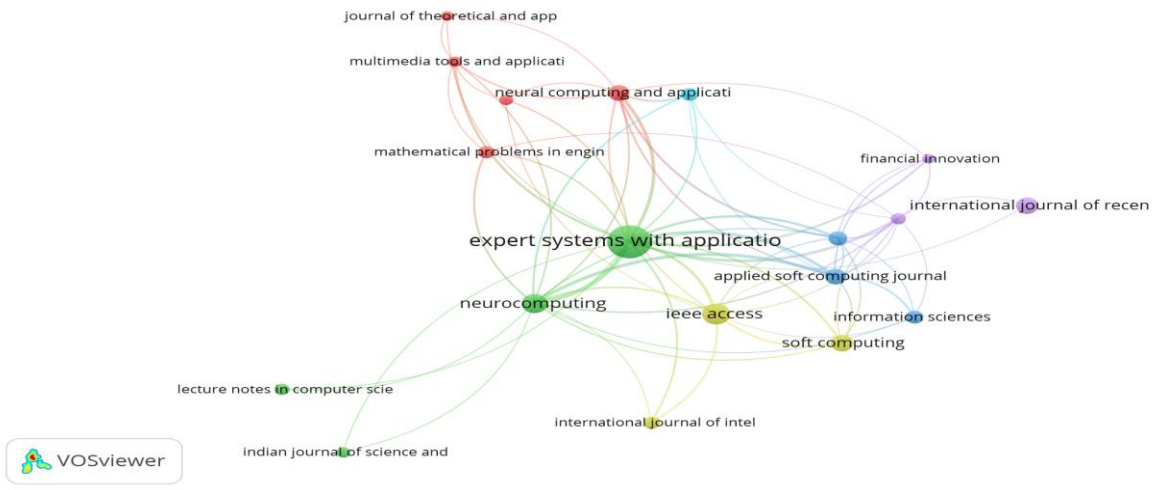

Figure-5: Network of top research outlets.

3.1.3. Most cited articles and Network analysis of article citation

Table 6 present the Most cited research articles. The article of Kim.K.J.(2003) is the most cited article with 1130 T.C. Following this, the article of Guresenet al. (2011), Chong et al.(2017), Ticknor J.L. (2013) andZhang & Wu (2009)are in descend with 484, 403,332,311 T.C. respectively.Other articles have less than 300 citations.Additionally, Table 6 provides detailed information about significant research articles, including their year of initial publication, age, total citations (T.C.), and associated links. Fig. 6 highlights how densely papers cite each other, revealing clusters of closely related research and the overall structure of the field. This density mapping techniques also show regions of high citation activity. The color gradients indicate citation density, with the denser areas in more intense colors.

Table 6- Most cited research articles

Author(s)	Pub. Year	Title	Age	Total Citations	TC/Age	Links
kim k.-j. (2003)	2003	"Financial time series forecasting using support vector machines"	20	1130	56.5	13
guresen e. (2011)	2011	"Using artificial neural network models in stock market index prediction"	12	484	40.33	8
chong e. (2017)	2017	"Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies"	6	403	67.17	14
ticknor j.l. (2013)	2013	"A Bayesian regularized artificial neural network for stock market forecasting"	10	332	33.2	10
yudong z. (2009)	2009	"Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network"	14	311	22.21	4
boyacioglu m.a. (2010)	2010	"An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange"	13	290	22.31	5

hadavandi e. (2010)	2010	"Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting"	13	284	21.85	4
dargan s. (2020)	2020	"A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning"	3	266	88.67	1
tsai c.-f. (2010)	2010	"Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches"	13	237	18.23	8
oliveira n. (2017)	2017	"The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices"	6	210	35	1
leigh w. (2002)	2002	"Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support"	21	206	9.81	6
hoseinzade e. (2019)	2019	"CNNpred: CNN-based stock market prediction using a diverse set of variables"	4	191	47.75	6
yu l. (2009)	2009	"Evolving least squares support vector machines for stock market trend mining"	14	177	12.64	2
xing f.z. (2018b)	2018	"Natural language based financial forecasting: a survey"	5	173	34.6	5
kim h.-j. (2007)	2007	"A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets"	16	158	9.88	9
singh r. (2017)	2017	"Stock prediction using deep learning"	6	149	24.83	3
kumar b.s. (2016)	2016	"A survey of the applications of text mining in financial domain"	7	147	21	2
majhi r. (2009)	2009	"Development and performance evaluation of FLANN based model for forecasting of stock markets"	14	141	10.07	6
kim k.-j. (2006)	2006	"Artificial neural networks with evolutionary instance selection for financial forecasting"	17	141	8.29	1
pang x. (2020)	2020	"An innovative neural network approach for stock market prediction"	3	138	46	1
kwon y.-k. (2007)	2007	"A hybrid neurogenetic approach for stock forecasting"	16	138	8.63	5
schumaker r.p. (2009a)	2009	"A quantitative stock prediction system based on financial news"	14	136	9.71	4
feng f. (2019)	2019	"Temporal relational ranking for stock prediction"	4	134	33.5	1
qian b. (2007)	2007	"Stock market prediction with multiple classifiers"	16	133	8.31	4
tanbeer s.k. (2009)	2009	"Sliding window-based frequent pattern mining over data streams"	14	131	9.36	1
huang c.-j. (2008)	2008	"Application of wrapper approach and composite classifier to the stock trend prediction"	15	131	8.73	5
wei l.-y. (2016)	2016	"A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting"	7	124	17.71	4

di persio l. (2016)	2016	"Artificial neural networks architectures for stock price prediction: Comparisons and applications"	7	120	17.14	2
asadi s. (2012)	2012	"Hybridization of evolutionary Levenberg-Marquardt neural networks and data pre-processing for stock market prediction"	11	119	10.82	4
picasso a. (2019)	2019	"Technical analysis and sentiment embeddings for market trend prediction"	4	118	29.5	3
li x. (2020)	2020	"Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong"	3	115	38.33	4
chung h. (2018)	2018	"Genetic algorithm-optimized long short-term memory network for stock market prediction"	5	114	22.8	6
shen l. (2004)	2004	"Applying rough sets to market timing decisions"	19	114	6	1
chang p.-c. (2009b)	2009	"A neural network with a case based dynamic window for stock trading prediction"	14	113	8.07	3
gepp a. (2018)	2018	"Big data techniques in auditing research and practice: Current trends and future opportunities"	5	112	22.4	1
qi m. (1999)	1999	"Nonlinear predictability of stock returns using financial and economic variables"	24	108	4.5	1
thawornwong s. (2004)	2004	"The adaptive selection of financial and economic variables for use with artificial neural networks"	19	107	5.63	5
chen y. (2007)	2007	"Flexible neural trees ensemble for stock index modeling"	16	105	6.56	1
nabipour m. (2020b)	2020	"Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis"	3	104	34.67	4
chang p.-c. (2009a)	2009	"Integrating a piecewise linear representation method and a neural network model for stock trading points prediction"	14	103	7.36	3
nabipour m. (2020a)	2020	"Deep learning for stock market prediction"	3	101	33.67	7
zhou x. (2018)	2018	"Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets"	5	100	20	2

Source: compiled from R-Bibliometrix and VOS viewer

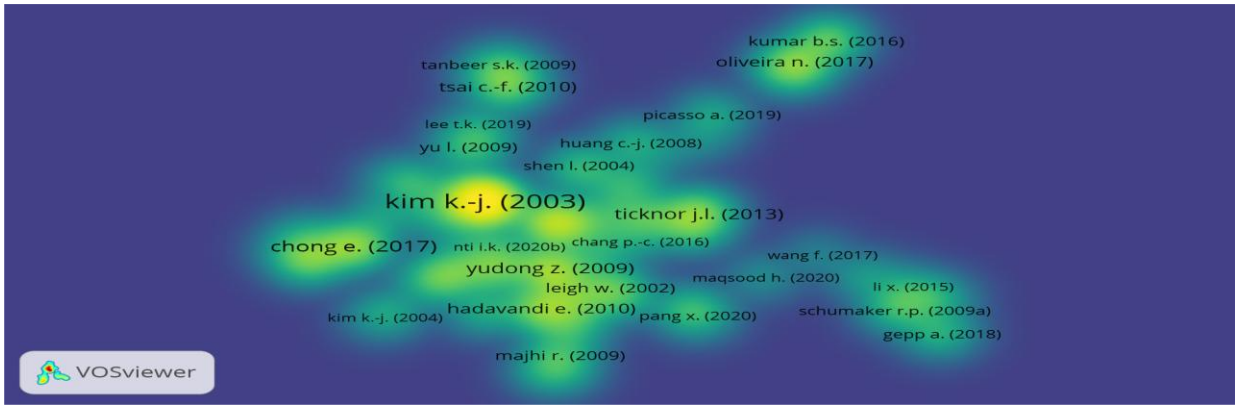

Fig. 6Network of Article Citations

Minnesota Journal of Business Law and Entrepreneurship

Volume 2026, No. 2

ISSN: 1540-3270

3.1.4. Influential countries:

Fig. 7 below illustrates the nations contributions and productivity on employing AI and MLmethods in stock market prediction field.The node sizes proportionally represent the number of published articles from each country, highlighting the magnitude of their research output in this domain.The highest number of publications are found in the pioneer countries, namely China, India and the US with 238,210 and 82 articlesrespectively on this topic.

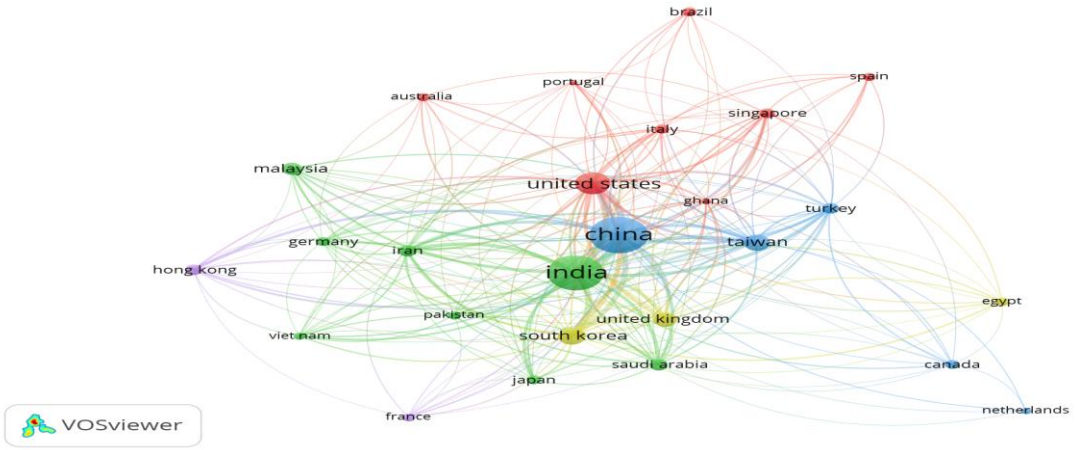

Fig. 7 Network of Top Influential Countries

Table -8: Top Influential Countries

Country	No of articles	T.C	A.C	TLS
China	238	4039	16.97	546
India	210	2461	11.72	577
United states	82	4020	49.02	363
South Korea	51	3516	68.94	321
Taiwan	47	1589	33.81	219
United Kingdom	37	1082	29.24	159
Malaysia	27	1204	44.59	65
Iran	23	1017	44.22	197
Saudi Arabia	22	371	16.86	119
Turkey	21	1079	51.38	148
Hong Kong	19	765	40.26	74
Singapore	18	829	46.06	86
Italy	15	420	28.00	54
Canada	14	260	18.57	24
Brazil	13	642	49.38	15
Germany	13	594	45.69	56
Japan	13	279	21.46	51
Australia	12	181	15.08	24
Pakistan	12	172	14.33	62

France	12	152	12.67	25
Spain	12	114	9.50	22
Egypt	11	147	13.36	33
Vietnam	10	254	25.40	61
Netherlands	7	212	30.29	8
Portugal	5	597	119.40	22
Ghana	5	193	38.60	65

Source: Compiled from R-Bibliometrix and VOS viewer

Articles can generally be categorized into two types based on author collaboration. In SCP(single country publications), a notable feature is the predominance of authors from a single country, indicating a high level of intra-country research collaboration. Conversely, MCP (multiple country publications) showcase a diverse range of authors from multiple countries, reflecting international or inter-country collaborations, highlighting the increasing globalization of scientific research.Fig.8 shows the status of research collaboration. The findings indicate thatChina has emerged as the leading contributor. India occupies second in terms of SCP but US advances to second in terms of MCP. The top 20 countries in collaborative research are as shown on Fig. 8.

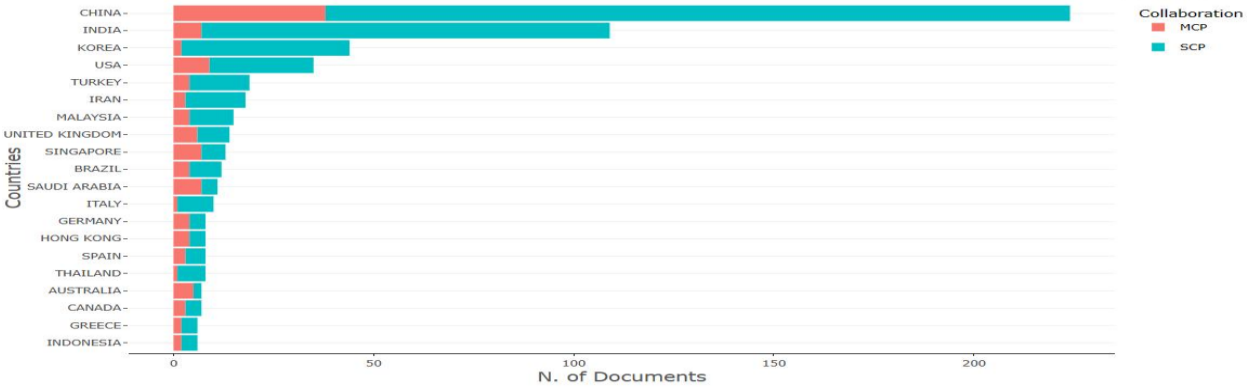

Figure-8: Corresponding Author's Countries

Countries' Collaboration World Map:

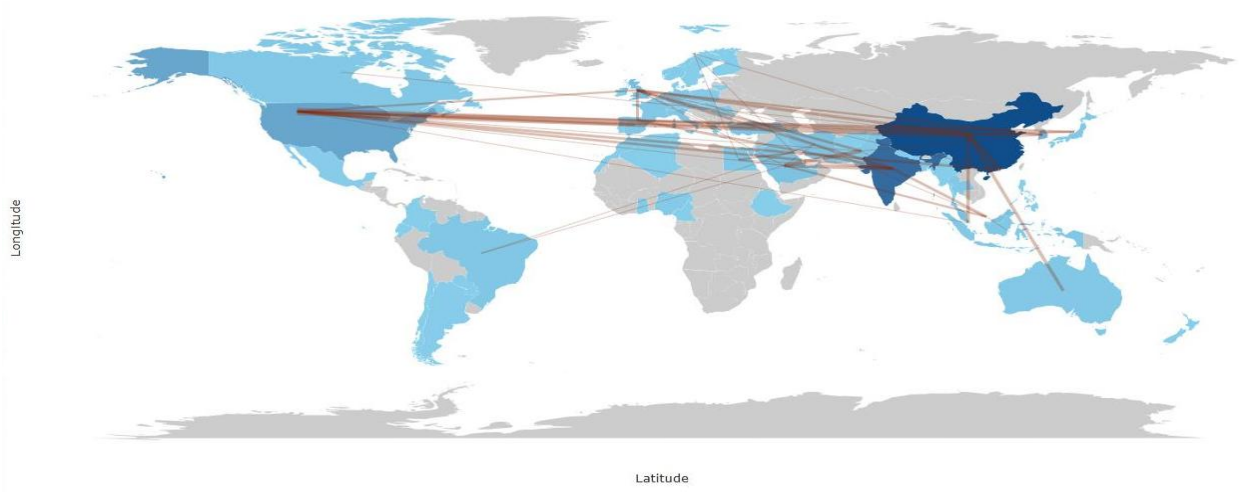

Figure- 9 Global Scientific Production and Collaboration Patterns

3.2 Conceptual structure: The third research question was addressed on investigating the potential gaps and interest areas in existing literature on employing AI and ML methods in stock market prediction field, quantitative text analysis was performed usinga Words Tree Map, Word Cloud, and co-word occurrence network analysis along with detailed co-word analysis tabulated with TLS at Table 9.

Also, to address the last research question on investigating the evolution of research topics on employing AI and ML methods in stock market prediction field, Words' Frequency over Time, Trend Topics Word Growth on a co-word map, Subject area distribution and Major areas of the literature was presented.

3.2.1 Quantitative Analysis of Word Frequency and Evolution in a Corpus of Text Data:

Keywords, which mirror the essence of research publications, play a vital role in locating relevant journals. As observed during extraction of bibliometric data, choosing correct words can streamline to achieve diverse research objectives. Fig. 10, Fig. 11 shows the Tree Map and Word-Cloud respectively of top 40 frequent words of research (author's keywords)employing AI and ML methods in stock market prediction field. Here to observe that 'stock market prediction'(127), 'stock prediction (124), 'deep learning'(99) and 'machine learning' (97) are the most frequently used by authors.

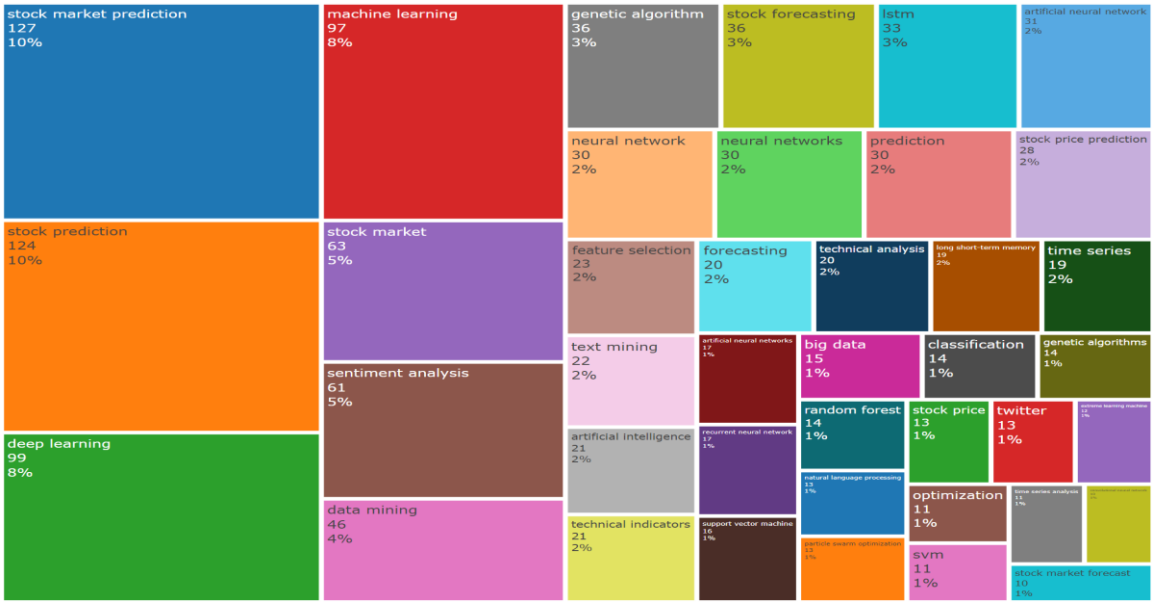

Figure-10 Tree Map showing the top 40 words

Figure-11 Word Cloud

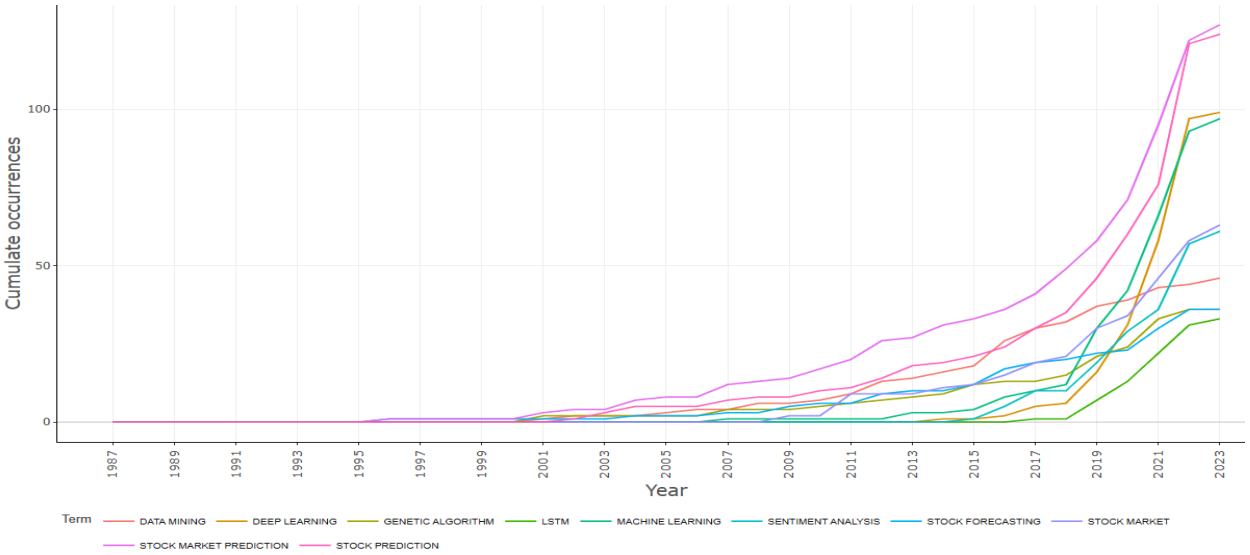

Figure-12 Frequency of Words over Time

3.2.2 Co-words analysis and C*o-occurrence network of keywords:*

Co-word analysis is a bibliometric technique to analyze the connectedness between words that co-occur in a set of documents. Co-words analysis involved examining the co-occurrence function with "Author's keywords" set as the unit. Fractional counting was selected for the counting option, with the default number of keywords (5) being utilized and 89 keywords met this threshold. However, when the default is modified to 10, only 41 keywords emerged. The 41 keywords along with TLS is appropriately displayed in Table 9.

Table-9 Co-word Analysis

S/N	*Keywords*	*Frequency*	*TLS*	*S/N*	*Keywords*	*Frequency*	*TLS*
1	*Stock Market Prediction*	*127*	*106*	*22*	*Long Short-Term Memory*	*19*	*19*
2	*Stock Prediction*	*124*	*96*	*23*	*Time Series*	*19*	*15*
3	*Deep Learning*	*99*	*95*	*24*	*Artificial Neural Networks*	*17*	*16*
4	*Machine Learning*	*97*	*87*	*25*	*Recurrent Neural Network*	*17*	*16*
5	*Stock Market*	*63*	*55*	*26*	*Support Vector Machine*	*16*	*15*
6	*Sentiment Analysis*	*61*	*59*	*27*	*Big Data*	*15*	*12*
7	*Data Mining*	*46*	*40*	*28*	*Classification*	*14*	*14*
8	*Genetic Algorithm*	*36*	*34*	*29*	*Genetic Algorithms*	*14*	*13*
9	*Stock Forecasting*	*36*	*29*	*30*	*Random Forest*	*14*	*13*
10	*Lstm*	*33*	*28*	*31*	*Natural Language Processing*	*13*	*13*
11	*Artificial Neural Network*	*31*	*28*	*32*	*Particle Swarm Optimization*	*13*	*10*
12	*Neural Network*	*30*	*27*	*33*	*Stock Price*	*13*	*10*
13	*Neural Networks*	*30*	*24*	*34*	*Twitter*	*13*	*13*

14	*Prediction*	*30*	*29*	*35*	*Extreme Learning Machine*	*12*	*10*
15	*Stock Price Prediction*	*28*	*24*	*36*	*Optimization*	*11*	*11*
16	*Feature Selection*	*23*	*20*	*37*	*Svm*	*11*	*8*
17	*text mining*	*22*	*22*	*38*	*time series analysis*	*11*	*11*
18	*artificial intelligence*	*21*	*16*	*39*	*convolutional neural network*	*10*	*9*
19	*technical indicators*	*21*	*21*	*40*	*stock market forecast*	*10*	*8*
20	*forecasting*	*20*	*20*	*41*	*stock price forecasting*	*10*	*9*
21	*technical analysis*	*20*	*20*				

Figure 13 illustrates visual representation of the frequency and strength of relationships between the set of authors keywords. Keywords closely related to each other are grouped into separate clusters. The size of each node reflects how frequently multiple authors have used a specific keyword.

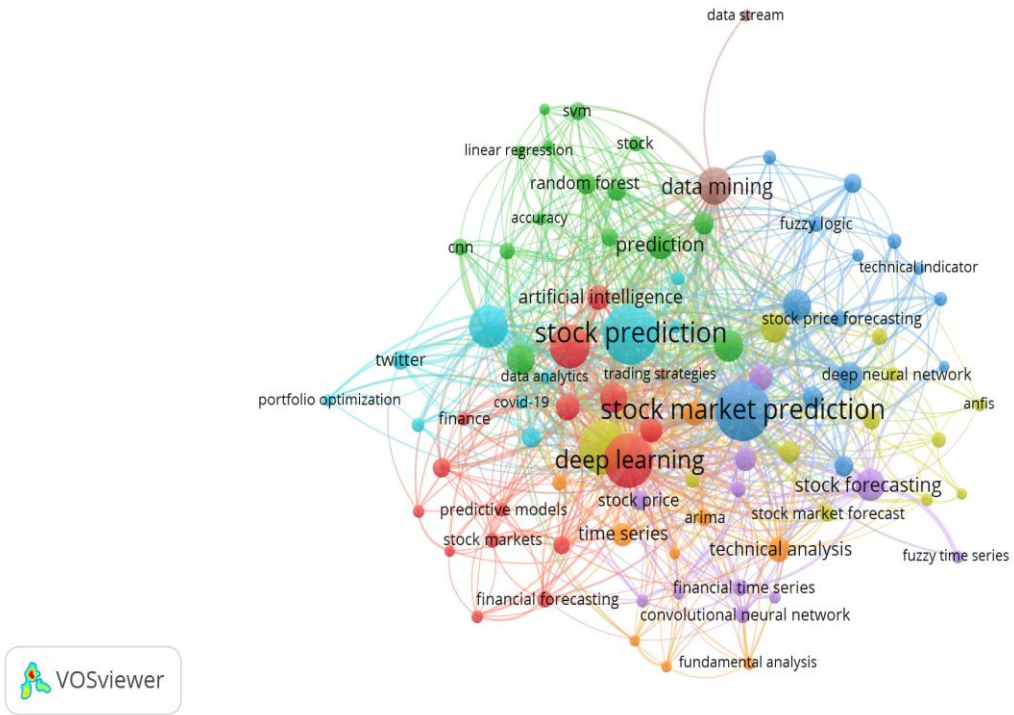

Figure- 13 Co-occurrence network of keywords

3.2.3. Subject area distribution:

It is crucial to analyze the roles of different disciplines and fields that have made contributions to this topic. According to a comprehensive search of Scopus, our analysis reveals that Computer Science is the most prominent discipline in this research area, accounting for 37.5% of all subject areas with 603 articles. Engineering and Mathematics follow with 310 and 195 articles respectively, while Business Management& Accounting and Decision Science also make significant contributions with 91 and 73 articles, respectively.The visualization in Figure 14 provides a breakdown of the relative contributions of the top disciplines in this field,

showcasing the significant influence of each area. Table 7 displays the publication frequency across the top eight disciplines, allowing for a deeper exploration of the research landscape.

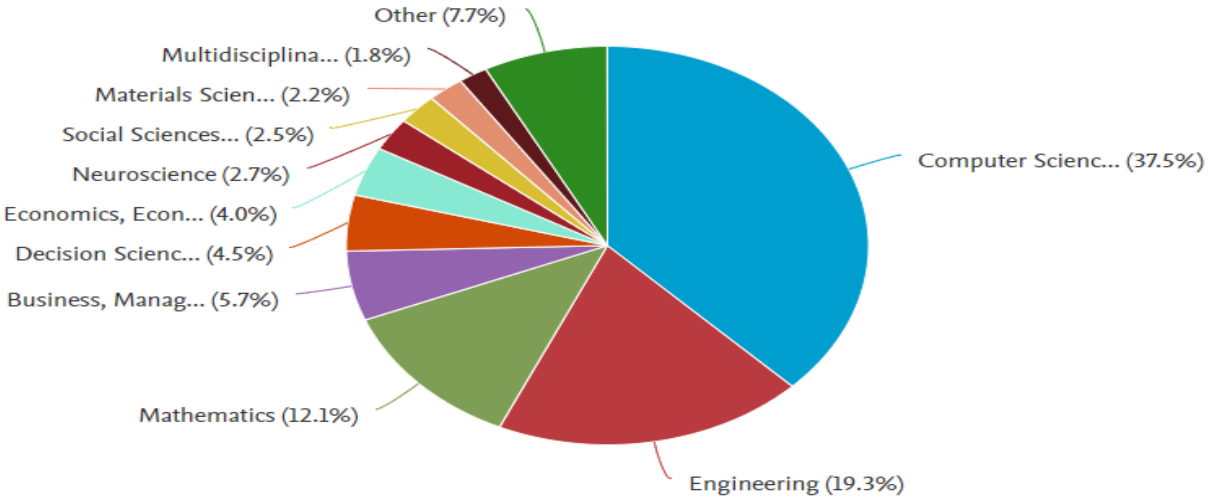

Figure- 14 Subject area distribution

Table 10 Disciplines participation:

Subject area	Articles Number
Computer Science	603
Engineering	310
Mathematics	195
Business, Management and Accounting	91
Decision Sciences	73
Economics, Econometrics and Finance	64
Neuroscience	43
Social Sciences	40

3.2.4. Major stock market research areas:

Many studies emphasize the importance of identifying keywords in understanding patterns and trends within a specific field. By analyzing key terms, researchers can gain a deeper understanding of the subject matter. This study expands author-keyword co-occurrence networks by incorporating a broader scope, extracting keywords from both titles and abstracts of 834 downloaded articles to provide more sophisticated understanding of the research environment.

We employed a retrieval strategy that involved setting a minimum threshold of 15 occurrences, as recommended by Yin et al. (2019) for conducting experiments to achieve ideal graphics.This threshold was determined using binary counting, considering a term relevant if it appears at least 15 times. Furthermore, to ensure high relevance, we also applied the standard benchmark of including 60% of the most relevant terms in the retrieved set. According to our co-occurrence analysis, key areas that require attention are depicted in Figure 15. A thorough examination of the titles and abstracts of the reviewed articles reveals the most frequently recurring keywords presented in Table 11.

Minnesota Journal of Business Law and Entrepreneurship

Volume 2026, No. 2

ISSN: 1540-3270

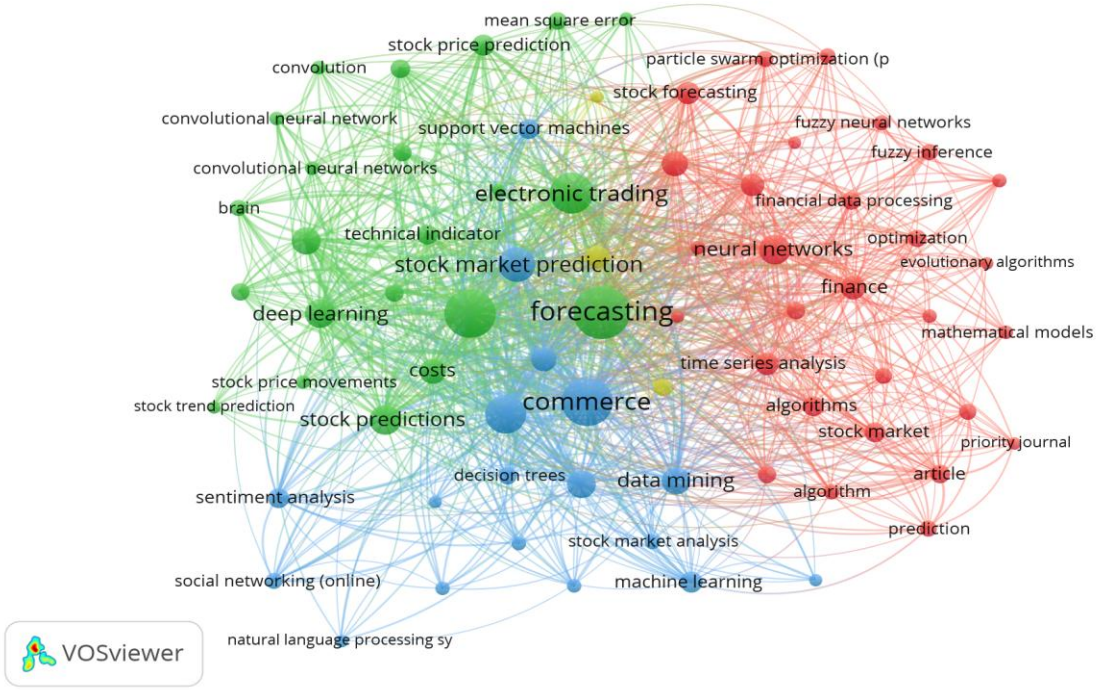

Figure-15 Three Clusters of Research Keywords

Table 11 Most frequently appearing words in titles and abstracts

Keywords	Occurrences	Avg. norm. citations
forecasting	377	1.3103
financial markets	300	1.4047
commerce	295	1.3439
electronic trading	202	1.4727
investments	175	1.2033
stock market prediction	139	1.6458
neural networks	102	1.488
stock predictions	100	1.1905
deep learning	92	1.9946
data mining	88	1.008
learning systems	86	1.6081
long short-term memory	84	1.1179
costs	69	1.5459
time series	67	1.5269
finance	61	1.2559
learning algorithms	59	1.6492
genetic algorithms	58	1.4394
stock price	56	1.0865
time series analysis	56	1.4097
stock forecasting	52	0.8767

stock price prediction	49	1.9052
machine learning	46	1.3074
algorithms	45	0.7229
technical indicator	43	1.2709
stock market	42	0.8062
sentiment analysis	40	1.7971
support vector machines	40	1.1425

4 Conclusion :

The stock market's role in the global economy is indispensableand we notice extensive application of machine learning methods in stock market analysis. However, there is limited focus towards comprehensive framework building that consolidates significantly important topics, areas and research field thereon. This study focus on addressing this gap by thoroughly investigating the literature published over the past few decades. A total of 834 scientific publicationson employing artificial intelligence and machine learning methods in stock market prediction field, are selected from the Scopus database during 1986-2023 for input.

China, India, and USare the top nations in count of articles published on employing AI and ML methods in stock market prediction. Specifically, these countries have collectively contributed 238,210 and 82 articles, respectively, demonstrating their significant advancements in this field. The top most journal in this domain is 'Expert Systems with Applications' (with 57 articles) and the most productive authors are LI X and WANG X (with 14 articles each) in Scopus database. The results show that 58 authors have published 68 documents individually, and 19.78% of authors have collaborated internationally.

The articles "Financial time series forecasting using support vector machines" by Kim, K. J. (2003) and "Using artificial neural network models in stock market index prediction" by Guresen et al. (2011) and "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies" by Chong et al. (2017) have received substantial recognition in this field with total citations 1130,484,403 respectively. Further, according to the results, China holds the top spot in terms of both SPC and MCP, making it the country with the highest contribution in academic publishing.The top contributing discipline is computer science, with 603 articles, representing 37.5% of all subject areas combined.

Our study, like most, has some constraints. Despite Scopus being one of the leading and most popular research databases among researchers, it inevitably excludes some significant publications that are not part of its collection. This exclusion limits the comprehensiveness of our analysis and introduces the potential for selection bias. Furthermore, the study could be broadened to encompass various types of documents and other databases. Despite the constraints, this review article aims to deliver a thorough, data-driven examination of the current state of machine learning applications in stock market prediction, offering insights into the most promising areas for future growth and development

References:

[1] Abdullah, M., Sulong, Z., &Chowdhury, M. A. F. (2024).Explainable deep learning model for stock price forecasting using textual analysis. *Expert Systems with Applications, 249*, 123740.https://doi.org/10.1016/j.eswa.2024.123740

[2] Arsenyan, J., &Piepenbrink, A. (2023). Artificial intelligence research in management: A computational literature review. *IEEE Transactions on Engineering Management, 71*, 5088-5100.https://doi.org/10.1109/TEM.2022.3229821

[3] Ayyildiz, N., &Iskenderoglu, O. (2024). How effective is machine learning in stock market predictions? *Heliyon, 10*(2).https://doi.org/10.1016/j.heliyon.2024.e24123

[4] Bansal, M., Goyal, A., &Choudhary, A. (2022).Stock market prediction with high accuracy using machine learning techniques. *Procedia Computer Science, 215*, 247-265.https://doi.org/10.1016/j.procs.2022.12.028

[5] Bello, I. T., Zhai, S., He, Q., Xu, Q., & Ni, M. (2021).Scientometric review of advancements in the development of high-performance cathode for low and intermediate temperature solid oxide fuel cells: Three decades in retrospect. *International Journal of Hydrogen Energy, 46*(52), 26518-26536.https://doi.org/10.1016/j.ijhydene.2021.05.134

[6] Cheng, L., Shadabfar, M., &SioofyKhoojine, A. (2023).A state-of-the-art review of probabilistic portfolio management for future stock markets. *Mathematics, 11*(5), 1148.https://doi.org/10.3390/math11051148

[7] Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications, 83*, 187-205.https://doi.org/10.1016/j.eswa.2017.04.030

[8] Dang, Q. V. (2019, December).Reinforcement learning in stock trading. In International conference on computer science, applied mathematics and applications (pp. 311-322). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-38364-0_28

[9] Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research, 133*, 285-296.https://doi.org/10.1016/j.jbusres.2021.04.070

[10] Ellegaard, O. (2018). The application of bibliometric analysis: disciplinary and user aspects. *Scientometrics, 116*(1), 181-202.https://doi.org/10.1007/s11192-018-2765-z

[11] Gaire, H. N. (2017). Forecasting NEPSE index: An ARIMA and GARCH approach. *NRB Economic Review, 29*(1), 54-68.http://dx.doi.org/10.3126/nrber.v29i1.52530

[12] Ghosh, B. P., Bhuiyan, M. S., Das, D., Nguyen, T. N., Jewel, R. M., Mia, M. T., ...&Shahid, R. (2024). Deep Learning in Stock Market Forecasting: Comparative Analysis of Neural Network Architectures Across NSE and NYSE. *Journal of Computer Science and Technology Studies, 6*(1), 68-75.https://doi.org/10.32996/jcsts.2024.6.1.8

[13] Guresen, E., Kayakutlu, G., &Daim, T. U. (2011).Using artificial neural network models in stock market index prediction. *Expert systems with Applications, 38*(8), 10389-10397.https://doi.org/10.1016/j.eswa.2011.02.068

[14] Hiransha, M. E. A. G., Gopalakrishnan, E. A., Menon, V. K., &Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia computer science, 132*, 1351-1362.https://doi.org/10.1016/j.procs.2018.05.050

[15] Jain, R., Vanzara, R., &Sarvakar, K. (2024, February). The Rise of AI and ML in Financial Technology: An In-depth Study of Trends and Challenges. In International Conference on Communications and Cyber Physical Engineering 2018 (pp. 329-341). Singapore: Springer Nature Singapore.https://doi.org/10.1007/978-981-99-7137-4_32

[16] Jrad, M. (2023). A role of artificial intelligence in the context of economy: Bibliometric analysis and systematic literature review. *International Journal of Membrane Science and Technology, 10*(3), 1563-86.https://doi.org/10.15379/ijmst.v10i3.1756

[17] Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing, 55*(1-2), 307-319.https://doi.org/10.1016/S0925-2312(03)00372-2

[18] Lim, W. M., & Kumar, S. (2024). Guidelines for interpreting the results of bibliometric analysis: A sensemaking approach. *Global Business and Organizational Excellence, 43*(2), 17-26.

[19] https://doi.org/10.1002/joe.22229

[20] Lin, Z. (2018). Modelling and forecasting the stock market volatility of SSE Composite Index using GARCH models. *Future Generation Computer Systems, 79*, 960-972.https://doi.org/10.1016/j.future.2017.08.033

[21] Mukherjee, S., Sadhukhan, B., Sarkar, N., Roy, D., & De, S. (2023). Stock market prediction using deep learning algorithms. *CAAI Transactions on Intelligence Technology, 8*(1), 82-94.https://doi.org/10.1049/cit2.12059

[22] Nassar, L., Okwuchi, I. E., Saad, M., Karray, F., &Ponnambalam, K. (2020, July). Deep learning based approach for fresh produce market price prediction. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7).IEEE. https://doi.org/10.1109/IJCNN48605.2020.9207537

[23] Rath, S., Das, N. R., &Pattanayak, B. K. (2024). An Analytic Review on Stock Market Price Prediction using Machine Learning and Deep Learning Techniques. *Recent Patents on Engineering, 18*(2), 88-104.https://doi.org/10.2174/1872212118666230303154251

[24] Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., &Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic reviews, 10*, 1-19.https://doi.org/10.1186/s13643-020-01542-z

[25] Saini, N., & Mittal, A. K. (2014). Forecasting volatility in indian stock market using State Space models. *Journal of Statistical and Econometric Methods, 3*(1), 115-136.https://www.scienpress.com/upload/jsem/vol%203_1_8.pdf

[26] Sheth, D., & Shah, M. (2023).Predicting stock market using machine learning: best and accurate way to know future stock prices. *International Journal of System Assurance Engineering and Management, 14*(1), 1-18.https://doi.org/10.1007/s13198-022-01811-1

[27] Sivadasan, E. T., MohanaSundaram, N., &Santhosh, R. (2024).Stock market forecasting using deep learning with long short-term memory and gated recurrent unit. *Soft Computing, 28*(4), 3267-3282.https://doi.org/10.1007/s00500-023-09606-7

[28] Sun, Z., Dong, W., Shi, H., Ma, H., Cheng, L., & Huang, Z. (2022). Comparing machine learning models and statistical models for predicting heart failure events: A systematic review and meta-analysis. *Frontiers in Cardiovascular Medicine, 9*, 812276.https://doi.org/10.3389/fcvm.2022.812276

[29] Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert systems with applications, 40*(14), 5501-5506.https://doi.org/10.1016/j.eswa.2013.04.013

[30] Vuong, P. H., Phu, L. H., Van Nguyen, T. H., Duy, L. N., Bao, P. T., & Trinh, T. D. (2024). A bibliometric literature review of stock price forecasting: from statistical model to deep learning approach. *Science Progress, 107*(1), 00368504241236557.https://doi.org/10.1177/00368504241236557

[31] Yin, X., Liu, H., Chen, Y., & Al-Hussein, M. (2019).Building information modelling for off-site construction: Review and future directions. *Automation in construction, 101*, 72-91.https://doi.org/10.1016/j.autcon.2019.01.010

[32] Zhang, J., & Lei, Y. (2022).Deep reinforcement learning for stock prediction. *scientific programming, 2022*(1), 5812546.https://doi.org/10.1155/2022/5812546

[33] Zhong, X., &Enke, D. (2019).Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial innovation, 5*(1), 1-20.https://doi.org/10.1186/s40854-019-0138-0

[34] Zou, J., Lou, J., Wang, B., & Liu, S. (2024). A novel deep reinforcement learning based automated stock trading system using cascaded lstm networks. *Expert Systems with Applications, 242*, 122801.https://doi.org/10.1016/j.eswa.2023.122801

[35] Zhang, Y., & Wu, L. (2009).Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert systems with applications, 36*(5), 8849-8854.https://doi.org/10.1016/j.eswa.2008.11.028