Machine learning and Deep learning Applications in Stock Price Prediction: Top Trend and Bibliometric Analysis

ArashSalehpour [1*], Karim Samadzamani [2]
[1]Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
[2]Depaetment of Computer Engineering , university Collage of Nabi Akram , Tabriz , Iran
Email:arash.salehpour@tabrizu.ac.ir(Corresponding Author), kszamani@ucna.ac.ir
Receive Date: 01 Dec 2023 Accept Date:20 April 2024

Abstract

This paper conducts an exhaustive bibliometric examination of Machine Learning and Deep Learning applications in Stock Price Prediction from 2013 to 2022. The stock market's economic significance and volatility underscore the interest in predictive models. Numerous studies use techniques like Support Vector Machines, Neural Networks, and Reinforcement Learning to forecast market trends. Complexities, including uncertainty and data density, pose challenges in financial forecasting. The study employs Scopus and advanced visualization tools for trend analysis. Queries focusing on "Deep learning," "Machine learning," and "Stock price prediction" yielded 131 papers from 93 sources. The analysis reveals a 36.55% annual growth rate, with China, India, and South Korea as leading contributors. The recurring and co-citation networks illuminate influential papers and prominent themes. "Automated news reading: Stock price prediction based on financial news using context-capturing features" emerges as a pivotal document. Thematic maps highlight evolving trends like "neural networks" and "financial markets." Applying Lotka's Law uncovers uneven author contributions. The paper emphasizes electronic trading, deep learning, and investment themes in stock price prediction research. This analysis offers insights into research trends, influential authors, and emerging themes. It is a valuable resource for researchers, practitioners, and policy-makers exploring the fusion of machine learning, deep learning, and stock price prediction.

Keywords: Stock Price Prediction, Bibliometric, Financial Markets, Deep Learning, Machine Learning

1. Introduction

One of the significant ways that publicly traded corporations may generate money is via the stock market, which is also the primary component of the capital market. The stock market is sometimes referred to as the "barometer of the economy," which indicates that its volatility is strongly tied to the expansion or contraction of the overall economic market. Stock market prediction has been an appealing issue for scholars in various professions. In particular, a significant number of research have been carried out to forecast the movement of the stock market by using machine learning techniques such as support vector machine (SVM), Neural Networks, Deep learning, and

reinforcement learning. Researchers from various professions have long been interested in predicting the stock market's direction. Financial markets have been intensively investigated for learning, a well-known method in many applications. Popular algorithms such as support vector machine (SVM) and reinforcement learning may assist in maximizing profit while keeping risk low when it comes to stock option purchases.[1, 2] The market for financial services is an intricate, ever-evolving, and non-linear dynamical system.[3] Because of these two primary considerations, bibliometric methods are currently applied to almost every area of research. Bibliometric metrics are used widely as indicators in various fields, including research performance, productivity, influence, and impact. This is the foundation upon which traditional bibliometrics is built and the area in which it excels best in its evaluative power. Financial forecasting is characterized by high uncertainty, a dense amount of data, noise, non-stationarity, a lack of organization, and hidden relationships. Politics, the status of the economy as a whole, and the expectations of traders are only a few factors that might influence the movement of prices in the financial markets. Consequently, predicting the changes in price in the financial market is a tricky endeavor. The results of academic investigations are increasingly pointing to the conclusion that price shifts in the market are not the result of random chance. Instead, they function in a manner that is very dynamic and non-linear, making it possible that the standard "random walk" assumption for futures prices is nothing more than a "cloak of randomness" covering a noisy non-linear process that's concealed behind it. [4-6]

One of the most common applications of machine learning methods is financial time series forecasting. Researchers have shown that advanced forecasting algorithms can effectively predict price fluctuations in the financial market. On the other hand, financial experts see the informational efficiency of the financial markets as a good indicator. Prediction of Prices in Financial Markets with the Application of Machine Learning Even though many financial economists are convinced that the EMH is accurate, a significant portion of the ML literature investigates the extent to which prices in financial markets can be predicted and the extent to which model-based trading is profitable. For instance, predicting financial markets is a common application area for the development of novel approaches and for demonstrating the potential of these methodologies. The dynamic regression modeling strategy, which predicts future market values based on previous changes in those prices and other price time series, has emerged as the method of choice in recent years..[7]The bibliometric analysis is a rigorous approach, using quantitative and statistical descriptions to characterize the evolving intricacies of knowledge in a particular topic.[8-10]

2. Data Collection

The Scopus database was searched, and R studio and biblioshiny software were used to analyze and visualize the trends. Data were obtained from the Scopus database, which is widely recognized as being suitable for bibliometric analysis of scientific publications due to its stringent evaluation process, comprehensive scientific publications, and influential and credible information provided. This is one of the reasons why the database has gained such widespread acceptance. The search results might be retrieved using the formula TS = ("Deep learning" OR "Machine learning" AND "Stock price prediction"). The document type was an article; English was the only language allowed, and the publication period lasted ten years.

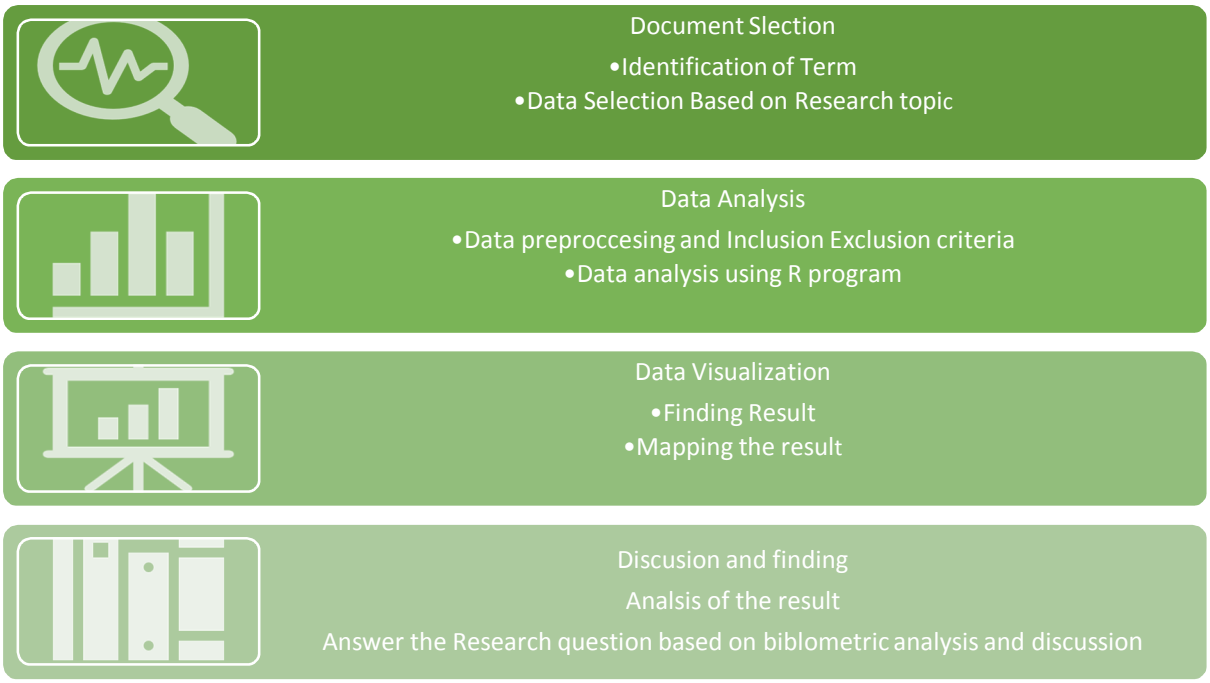

Fig.1.Data Analysis Process

1.1.1 Loading and converting data

The second step, known as "loading and converting data," comprises changing the data into an appropriate format and analyzing it using bibliometric tools. This procedure is known as "loading and converting data." The next step is to make the data seem to be of more outstanding quality. Bibliometrics data may be accepted chiefly; however, cited references may contain several copies of the same book published by different authors because the great majority of bibliometrics data may be depended on. The precision of the underlying data determines the calculation's precision. For example, it is possible to detect duplicate papers and misspelled phrases using different preprocessing techniques. A method for choosing and fine-tuning keywords, making a query based on both the TITLE, KEYWORD and the

ABSTRACT, the search was done., as shown in Table 1.

1.1.2 Data Analysis Tools

Quantitative data analysis is turned on its head by bibliometrics. Using quantitative methods, Bibliometrics studies scholarly publications, such as journal articles, and the citations that support them. We use Biblioshiny as the primary tool with R studio. [3]Microsoft Office Excel (2019). (Microsoft Corporation, Redmond, Washington, USA) was used to manage and analyze the characteristics of the publications. It visually analyzed keywords and the author's contributions and countries. The clustering algorithm creates a co-citation network and performs density visualization analysis.

1.1.3 Word Cloud

The most frequent word all over the documents is “forecasting, “which appears 76 times; “financial markets” appears 66, “stock price prediction” 54, “commerce”36, electronic trading 36, “deep learning” 35, and “long short-term memory”28, “investments”27 “costs”25, machine learning 23.

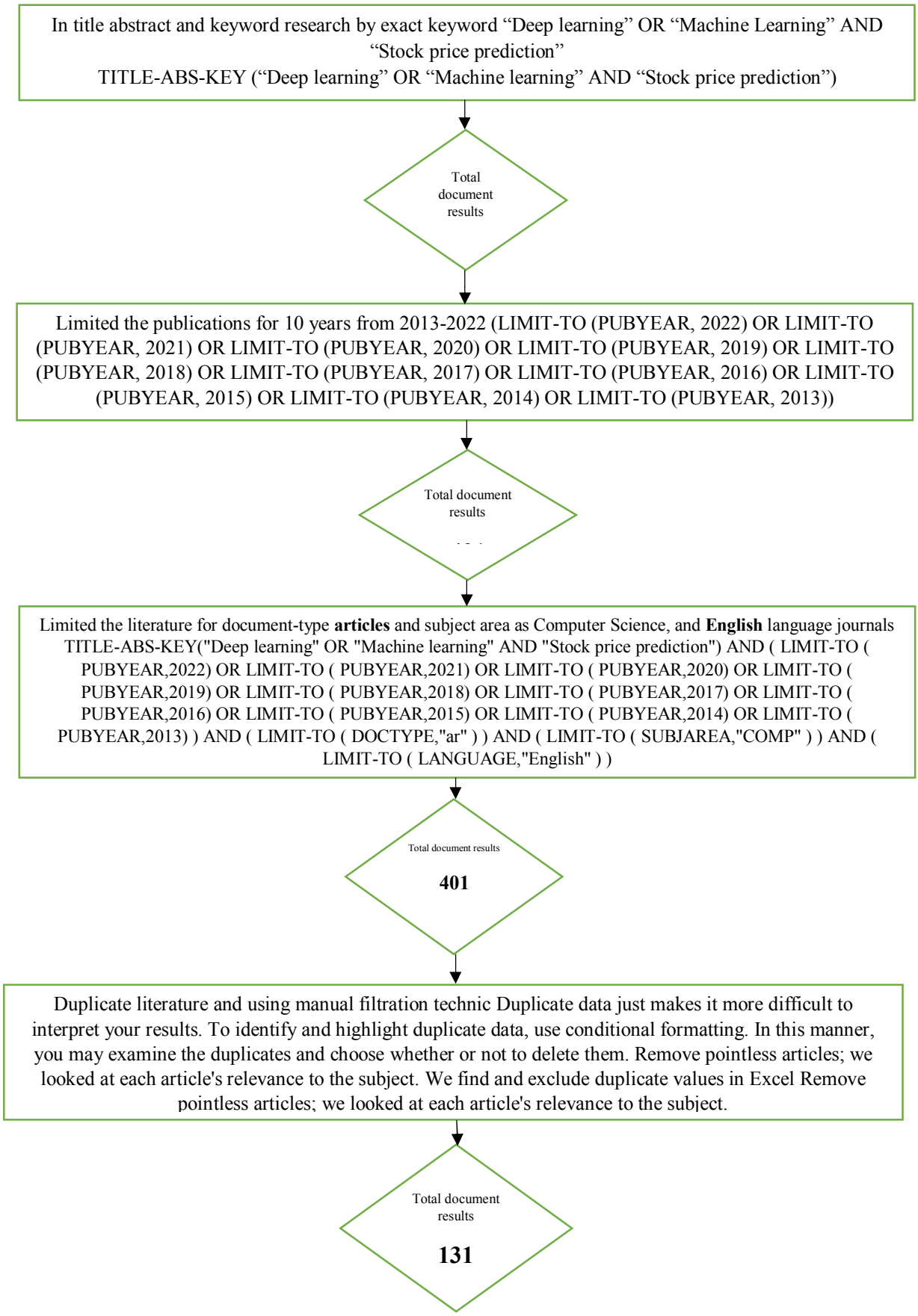

Fig.2.Step-by-step query and refin*e*

Table 1. Step-by-step query and refine

Final query used for data collection.
TITLE-ABS-KEY (("Deep learning" OR "Machine learning" AND "Stock price prediction")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))

1.2 Study approach and tools

Our database contains 131 papers from 93 sources, and the collection's period was 2013-0222. The average number of citations per document was 12.72, and we identified 381 authors among 131 documents. The annual growth rate is 36.55%, and international authorship is 12%. We use biblioshiny in this section and RStudio.

Table 2. Main information of Data The main Information and tables illustrate the distribution of literature from 2013-2022, achieving an average growth rate of 36.55.%. It is noticeable that research productivity was almost stable from 2013-2018, but considerable

Description	*Results*
MAIN INFORMATION ABOUT DATA	
Timespan	2013:2022
Sources (Journals, Books, etc)	93
Documents	134
Annual Growth Rate %	36.55
Document Average Age	1.72
Average citations per doc	12.72
References	4617
DOCUMENT CONTENTS	
Keywords Plus (ID)	591
Author's Keywords (DE)	368
AUTHORS	
Authors	381
Authors of single-authored docs	7
AUTHORS COLLABORATION	
Single-authored docs	7
Co-Authors per Doc	3.07
International co-authorships %	12.69
DOCUMENT TYPES	
	3
article	131

1.2.1 Annual scientific production

We can conclude that the highest number of publications related to Stock price prediction with machine learning and deep learning focus in the year 2021 .

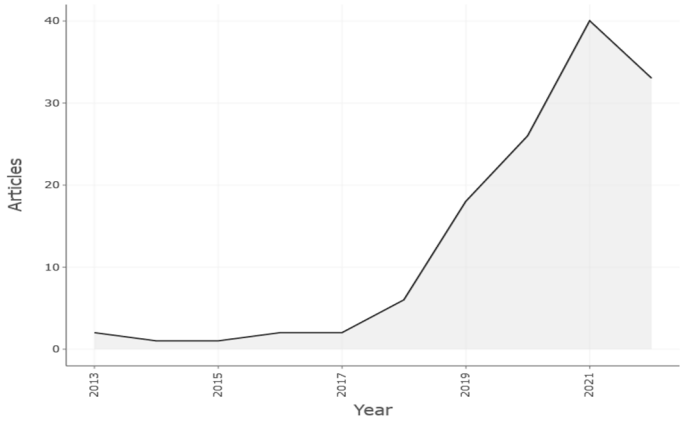

Fig.3. Annual Growth Rate

1.2.2 The countries that publish the most applications of Machine learning and Deep learning in Stock Price Prediction are based on our survey.

The following table presents a ranking of the countries in terms of their level of participation in research in stock price prediction using machine learning and deep learning. This investigation was based on calculating the writers' associations with their own countries, and the findings revealed that China was the most prolific and influential country.

Table 3. The countries that publish the most

region	*Freq*
CHINA	113
INDIA	101
SOUTH KOREA	34
AUSTRALIA	22
UK	15
USA	13
INDONESIA	12
EGYPT	10
BANGLADESH	8
ITALY	6

1.2.3 Most cited countries

When compared to nations located all across the world, China stands out as having the highest total number of 527 citation

.

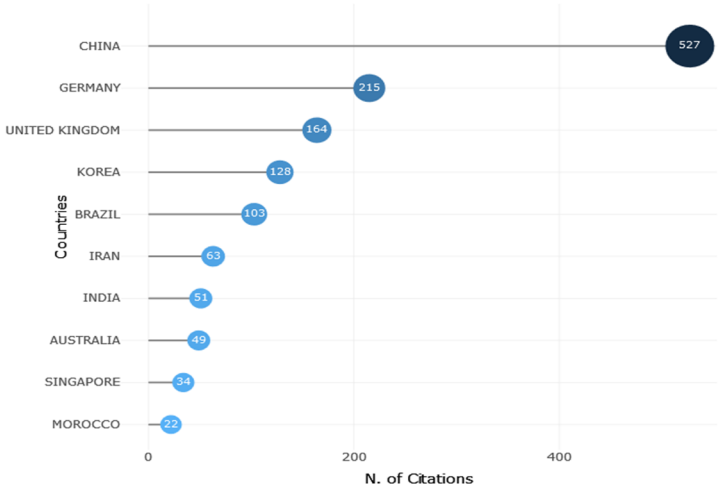

Fig.4 Most cited countries

1.2.4 The distribution of research publications

In addition, Figure 7 illustrates the distribution of research publications on Machine learning and Deep learning applications in Stock Price Prediction around the globe. A significant amount of research has been conducted and published on the subject of Stock Price Prediction. The volume of publications from each nation has caused their countries to become represented in the color patterns. Figure 7 presents a ranking of the ten countries that have contributed the most to the total number of scientific articles. With 113 document pages, China is the country with the most production. India is just a little behind China in second place with 100 papers, while South Korea comes in second with 34 papers. Compared to other nations, Australia holds the fourth position, with 22 papers covered. The next place goes to the United Kingdom with 15 pieces of paper, followed by the United States with 13 sheets, and Indonesia with 12 papers.

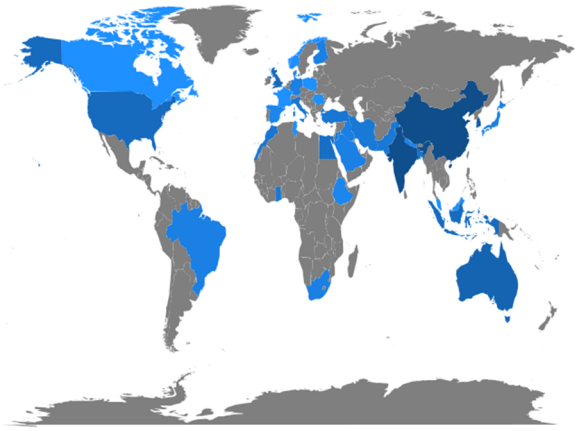

Fig.5. The distribution of research publications

1.2.5 Corresponding Author's Country

Figure 8 further elaborates on the related author's nationality, with the resulting graph revealing China, Australia, and India as among the nations with the most significant MCP counts of scientific papers addressing Machine learning and Deep learning applications in Stock Price Prediction. In this analysis, China, India, and South Korea considered SCP (single-country production). These nations have significantly impacted Machine learning and Deep learning applications in Stock Price Prediction.

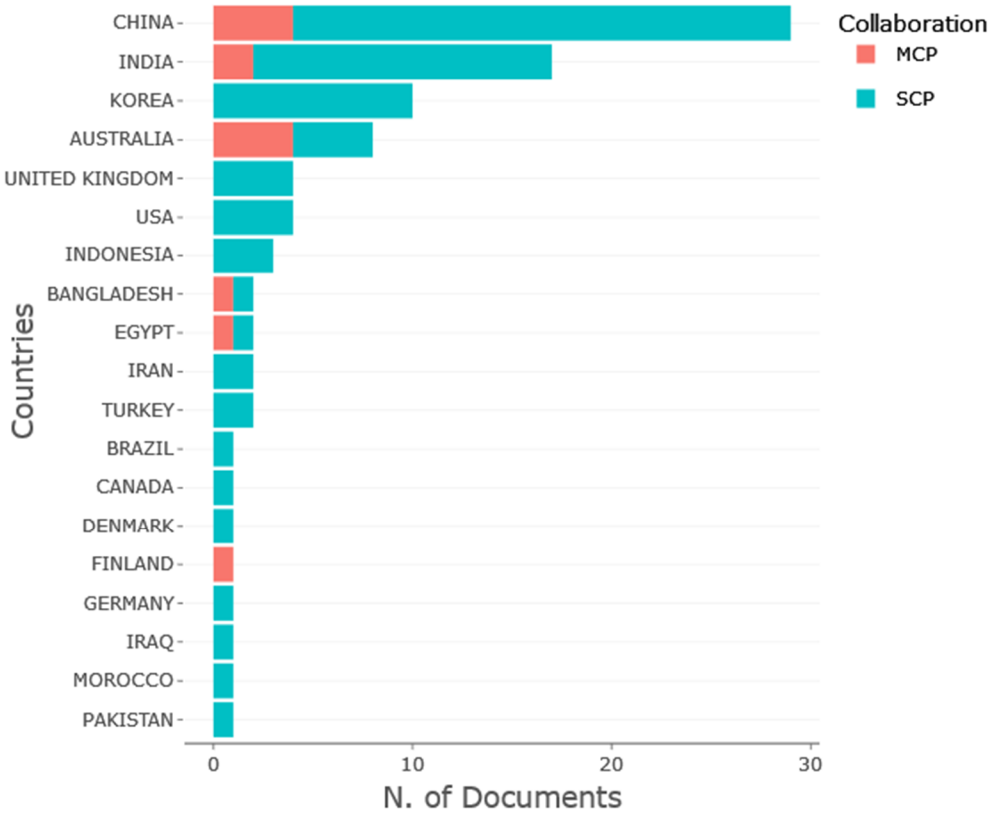

Fig.6. The Most Corresponding Author's Country

1.2.6 Countries' Production over Time

Different nations and economies make different contributions to the ever-increasing amount of published material that exists in the globe. When taken as a proportion of the overall world production, the publishing output of any nation or economy reveals that the historically significant producers have contributed the most. In the high-income economies (China, Germany, and other similar countries), the most increased publication occurred in 2020. China's growth rate is notable.

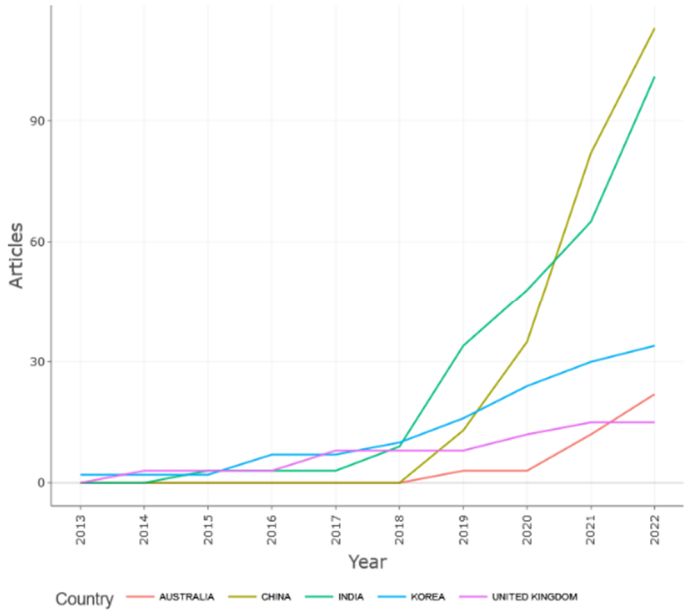

Fig.7. Countries' Production over Time

1.2.7 Most Relevant Affiliations

The data came from articles that were published in Scopus during the years 2013 and 2022. Because it identifies the institution in the region that is doing the most remarkable research activity, this data may be used by funding agencies and early-career researchers. Researchers may find this material helpful in locating a trustworthy source in this area and acquiring articles pertinent to their work. The distribution of scientific documents by institutions or connections is broken out and explained in Figure 10. According to the findings, ten different universities have produced the most work on the subject of research. The university that came in first position is located in China and is called Nanjing Tech University. It is the premier institution in producing scientific papers on stock price predictions. The university has contributed 8 document articles. The Capital University of Economics and Business, located in China, is the second most represented affiliation in producing scientific materials. In addition, there are educational establishments such as the Chinese Academy of Science in China. These universities are examples of establishments that have produced six publications. Each of the following universities: Symbiosis International University in India 6 publications; Dali University in China produced five publications; Seoul National University Soongsil University, both in South Korea and Sri Sivasubramanian Nadar College of Engineering in India; and Swinburne University of Technology located in Australia each of which has five publications, are the most recent universities to publish papers in Machine learning and Deep learning applications in Stock Price Prediction.

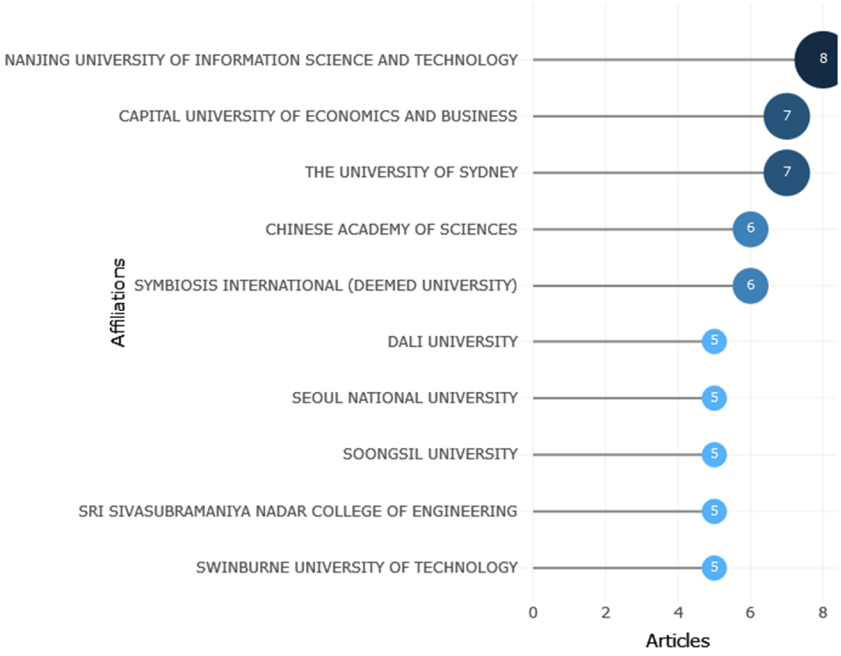

Fig.8. Affiliations

1.3 Citation Analysis

The average citation per document is 15.37

1.3.1 Most Local Cited Sources (from Reference Lists)

Over the last several years, emphasis has shifted toward using the number of citations received as a proxy for academic production. The primary results indicated a substantial growth in intellectual creation. Automated news reading: Stock price prediction based on financial news using context-capturing features is the most globally cited document in this study; to see whether textual information from financial news articles can be used to increase stock price prediction accuracy, we look at past techniques that only produced accuracy close to random guessing chance. We augment current text mining approaches by utilizing more expressive features to represent text and by incorporating market input into our feature selection process. As shown in this paper, classification accuracy may be improved by using a robust feature selection in conjunction with complicated feature types. Thus, over-fitting is avoided when using a machine learning methodology because of our method's ability to choose semantically meaningful features. In addition, we show that our strategy is very successful in the real world of trading. The approach may be used in any other application field as long as you have textual information and accompanying impact data.[11]The most locally cited source is Expert system with Application, with 141 citations, IEEE Access with 87, then system with Application 79 Neurocomputing 72 Expert System with Application 52 Plos One 31 Applied Soft Computing30 Procedia Computer Science 30 The Journal of finance29decision Support System 20

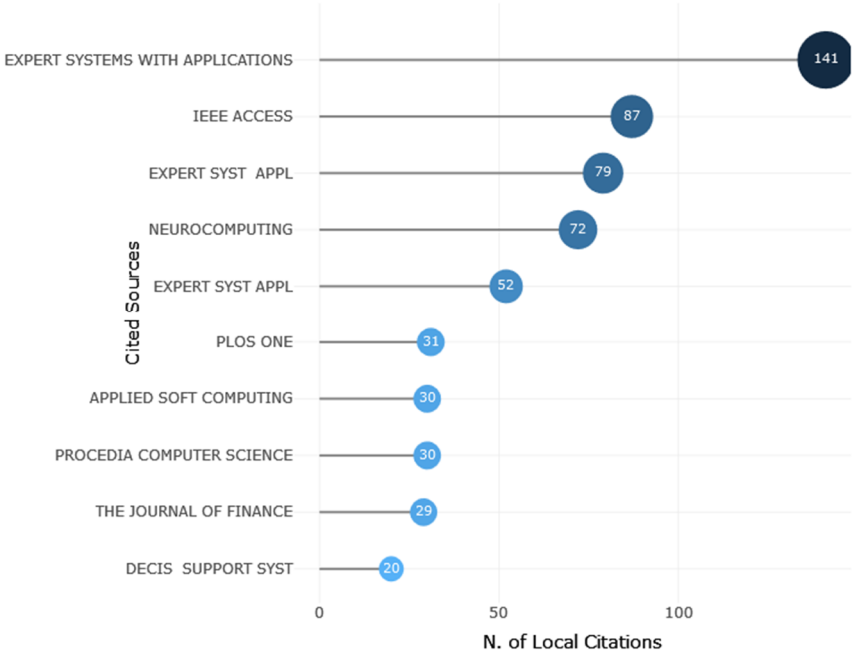

Fig.9.most cited sources

1.3.2 Source clustering through Bradford's Law

In physics, Bradford's Law of Scattering governs scattering and declining returns. As Bradford said when he first proposed his concept in 1948, "There are a few extremely prolific magazines, a greater number of more middling producers, and a yet larger number of steadily falling output" for each field of study. A Pareto distribution describes this shape in several fields. Consider the scenario when a researcher has identified five essential scientific periodicals for his or her field. Let us assume that there are 12 relevant papers published in such journals every month. Now imagine that the researcher must consult an extra ten journals to locate another twelve relevant papers. If so, the researcher has a Bradford multiplier bm of 2 (10/5). This researcher will have to scour times as many publications as possible to find each additional dozen papers. Most researchers recognize no value in continuing their search after reviewing the contents of the first five, ten, twenty, forty, etc. journals. The journals that are most frequently cited in the research that has been done on a particular issue or subject area are those that fall into the "top third" (also known as the "core" of the journals for that issue or subject area). As a result, these journals are the most likely to be of interest to academics who work in that particular field. The quantity of citations received by journals in Zone 2 is considered average. Journals in Zone 3, however, make up the "long tail," which is seldom cited and is hence seen as having little relevance to the subject at hand.[12]Bradford's law is used in Fig. 12. Bradford's law describes the distribution of papers in a field throughout a set of journals, with the bulk of articles clustering around a small set of authoritative

publications. The findings demonstrate a high percentage of journals that are released with a limited quantity of documents. Bradford's law said, as seen in Figure 12, that the leading journals that have been published in Stock Price Prediction, such as Expert System with Application, International Journal of Recent Technology and Engineering, Applied Intelligence, and IEEE Access, are the leading journals in Stock Price Prediction using machine learning or deep leaning.

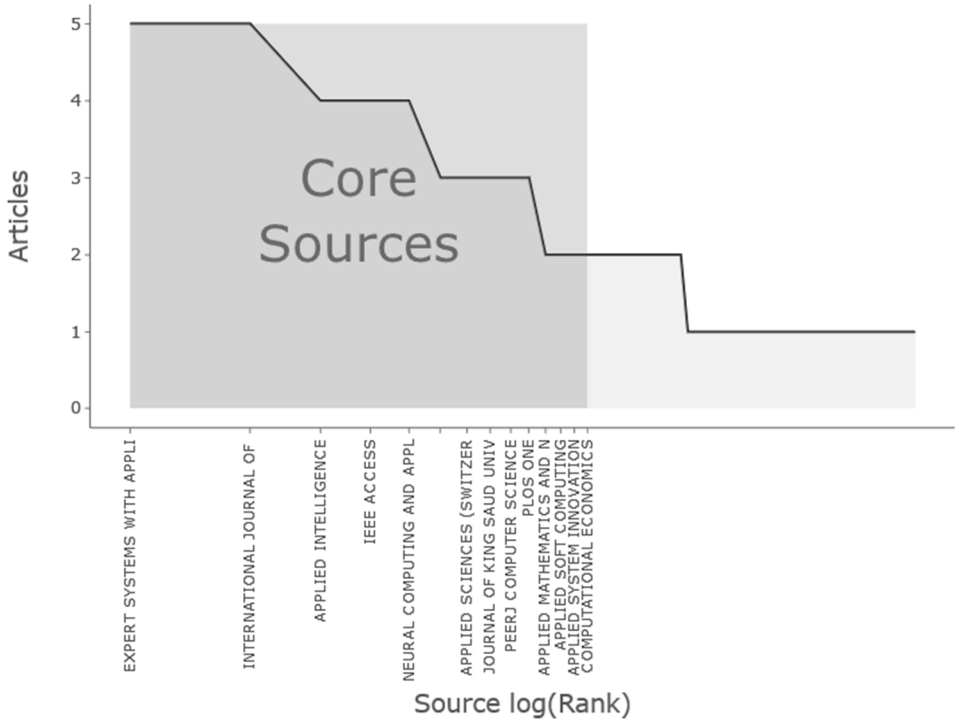

Fig.10. Source clustering through Bradford's Law

1.4 Network Visualization

1.4.1 Lotka's Law and productivity patterns of authors

The application of Lotka's Law of Scientific Production offers a framework for investigating inequalities in the patterns of authors' productivity in a particular subject and throughout a specific period. Since its release, several writers have used Lotka's Law within the context of the written works of various academic fields. The implementation of Lotka's Law of Authorship Productivity in the subject of Dentistry Literature is fruitful and productive. The equation below summarizes Lotka's Law as:

$$X^{\alpha}Y = K \quad [13]$$

where Y is the number of authors producing X number of articles in any given subject area [13]To rephrase, Lotka's rule states that for every x writer, there are about 1/x authors who contribute just once, and 60% of all authors contribute only once. It follows that around 60% of writers in any given subject will have just one publication, 15% will have two publications (1/22 times 60), 7% will have three publications (1/32 times 60), and only approximately 6% will contribute up

to 10 contributions apiece in the literature of any discipline. Based on Figure 13, we can conclude the findings of the research, which had a large number of contributors. Of these contributors, about 281 writers contributed just one document, which accounts for 0.953 of the total contributions made by the authors.

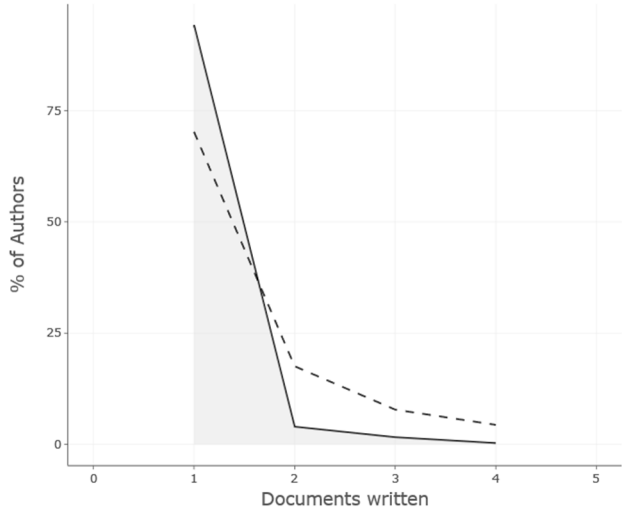

Fig.11. Author Productivity through Lotka's Law

1.4.2 Most Global Cited Documents

Top 10 most locally cited Documents top two are "Automated news reading: Stock price prediction based on financial news using context-capturing features"in Decision Support Systems [11]With a total citation of 215, a TC per year of 21.54, and a normalized TC of 1.76. The second document, titled "Deep learning-based feature engineering for stock price movement prediction" Knowledge-Based Systems [14] with total citation of 155 and TC per year 38.75 and normalize TC 6.37

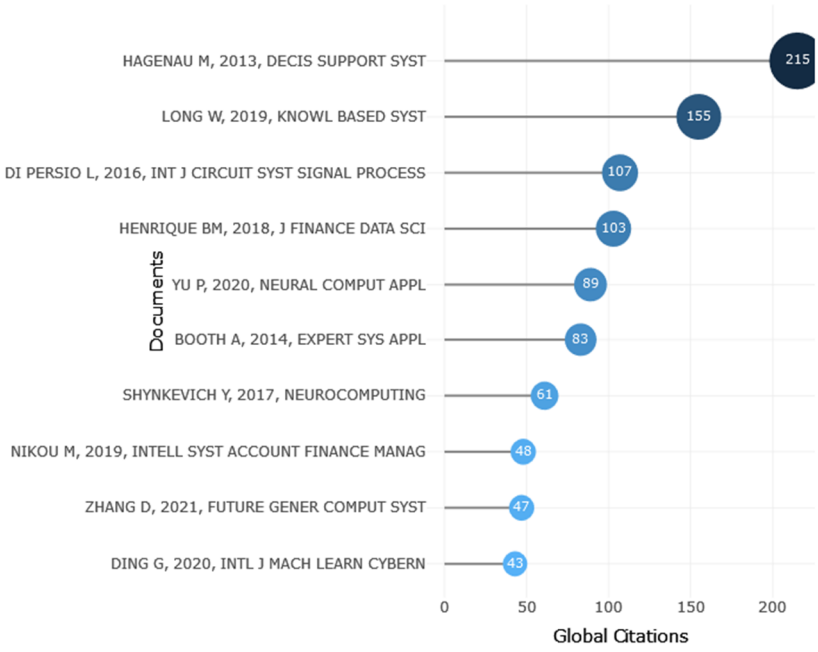

Fig.12. Most Global Cited Documents

1.4.3 Most Local Cited References

Top 10 most locally cited References in the literature. The most cited paper introduces Recurrent backpropagation as a slow learning method to retain knowledge over lengthy periods, mostly because of inadequate, declining, incorrect backflow. Propose a new, efficient gradient-based solution we call long short-term memory to solve it (LSTM).[15]

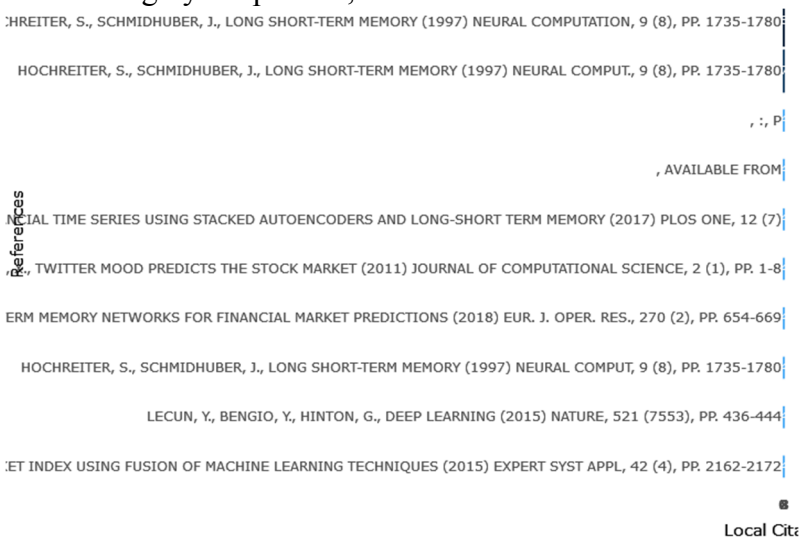

Fig.13.Most Local Cited References

1.5 Co-occurrence Network

A co-occurrence network is an undirected graph constructed using a corpus of documents as its starting point. Each node in the graph represents a unique word in a vocabulary, and each edge represents the frequency with which two words co-occur in a document. In order to show the links between words in a corpus of texts and to extract information about those associations, use co-occurrence networks; for instance, you might use a co-occurrence network to figure 17 out the terms that are most often found along with the specific word.

Table 4. Table of Co-occurrence Network

Node	*Cluster*	*Betweenness*	*Closeness*	*PageRank*
algorithm	1	0.802585	0.011905	0.011671
algorithms	1	0.062457	0.011494	0.009035
investment	1	1.014941	0.012048	0.012199
article	1	0	0.011236	0.008442
neural networks, computer	1	0.165673	0.011494	0.009054
prediction	1	0	0.011236	0.008442
forecasting	2	239.3254	0.020408	0.101109
financial markets	2	173.9004	0.019608	0.095172
stock price prediction	2	119.5112	0.019231	0.080455
commerce	2	82.28423	0.018519	0.058166

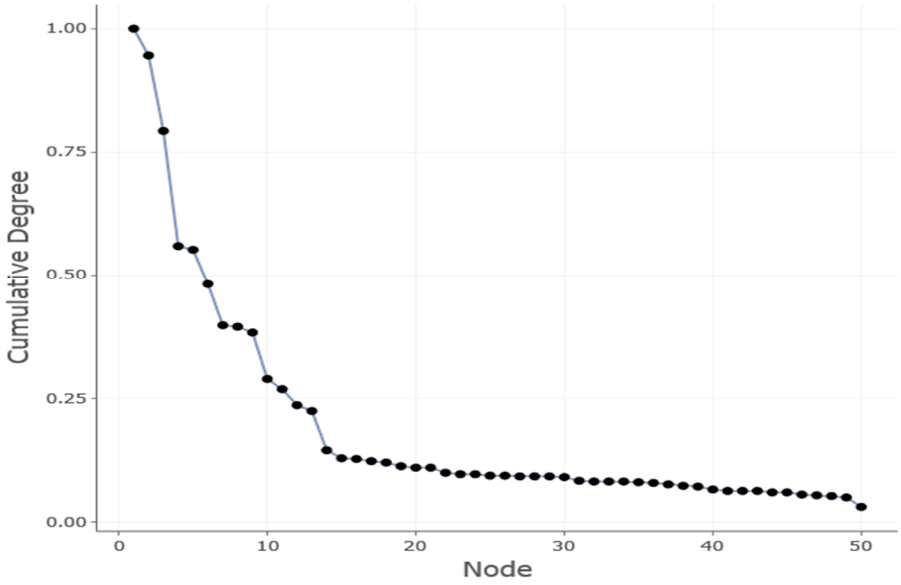

Fig.14. Degree Plot of Co-occurrence Network

Figu.15. Co-occurrence Network

1.6 Co-citation Network

With a field of papers, the first two articles include Rochester's. 1997 with Betweenness of 496, through the development of a cutting-edge, highly effective gradient-based technique named long short-term memory (LSTM). By truncating the gradient when it does not cause damage, LSTM can learn to bridge minimum temporal gaps of more than 1000 discrete-time steps by imposing continuous error flow through constant error carousels inside specialized units. Fischer t. 2018 with a Between's of 180.

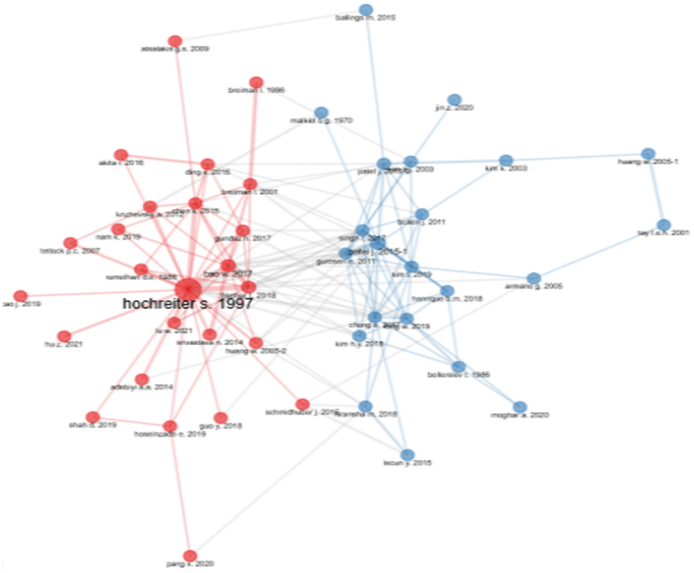

Fig.16. Co-citation Network

1.7 Word Dynamics

We discovered that research on stock market prediction expanded with a rise in themes, such as electronic trading and machine learning subjects in 2019.

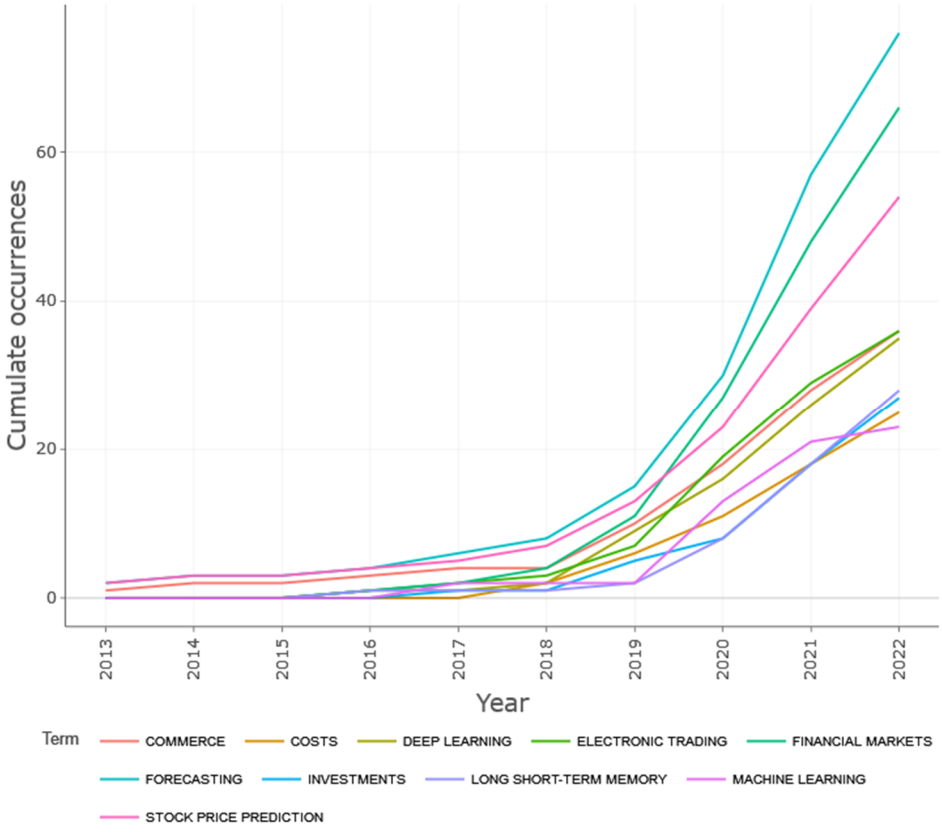

Fig.17. Word Dynamics

1.8 Trend Topics

Bibliometric analysis of the academic literature in terms of the subject distributions of the academic material This investigation makes it possible to get information on the publishing patterns of authors who acknowledge that their work is a part of the bibliometric study and the phrases they use in their work.

Figure 20 demonstrates what has been hot in Machine learning and Deep learning applications in Stock Price Prediction. Each year has a unique set of themes to explore. From 2018 through 2021, the terms "neural networks" and "stock price movements" appeared in 12 different contexts, making it one of the trendiest themes in writing. The writers discussed the second issue in this trend, "Commerce," " Machine Learning," and electronic trading," 95 times between 2019 and 2021. Moreover, the word "stock price prediction," "financial markets," and "forecasting" was used 195 times to describe the subject of study between 2020 through 2022. Nonetheless, neural networks and Stock prices are some of the trend issues the writers explore in 2022.

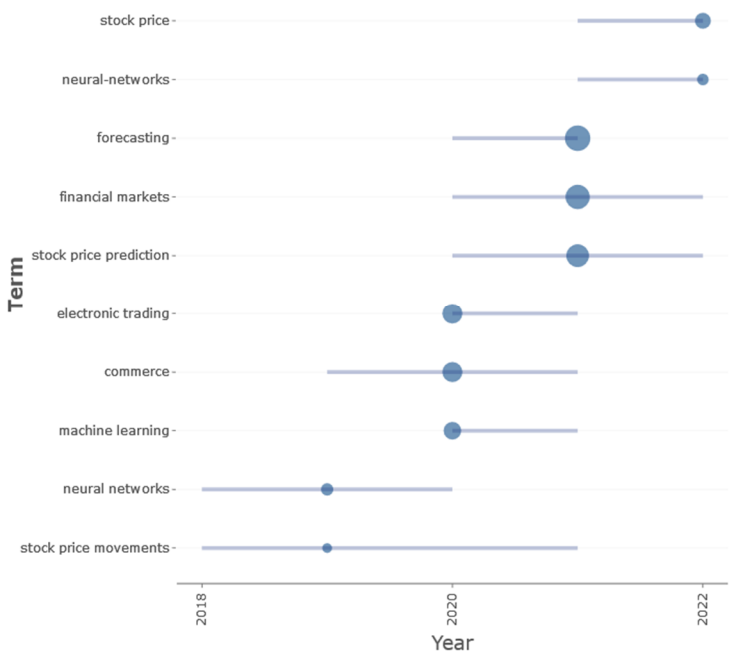

Fig.18. Trend Topic

1.9 Thematic Map filed keywrods plus

A "thematic map" is a kind of map that zeroes down on a single overarching concept or topic as its primary point of interest. How a specific topic or distribution differs around the world is highlighted by a thematic map. Bibliometrics use the theme map to sketch out the conceptual structure of a topic. The latter uses a co-occurrence network of phrases to evaluate the dominant topics, themes, and trends in academic discourse in a specific discipline.

Table 5. Table of thematic map top ten records

Occurrences	*Words*	*Cluster*	*Cluster_Label*
5	algorithm	1	algorithm
5	algorithms	1	algorithm
5	investment	1	algorithm
4	article	1	algorithm
4	neural networks, computer	1	algorithm
4	prediction	1	algorithm
2	marketing	1	algorithm
2	economic aspect	1	algorithm
2	human	1	algorithm
2	humans	1	algorithm

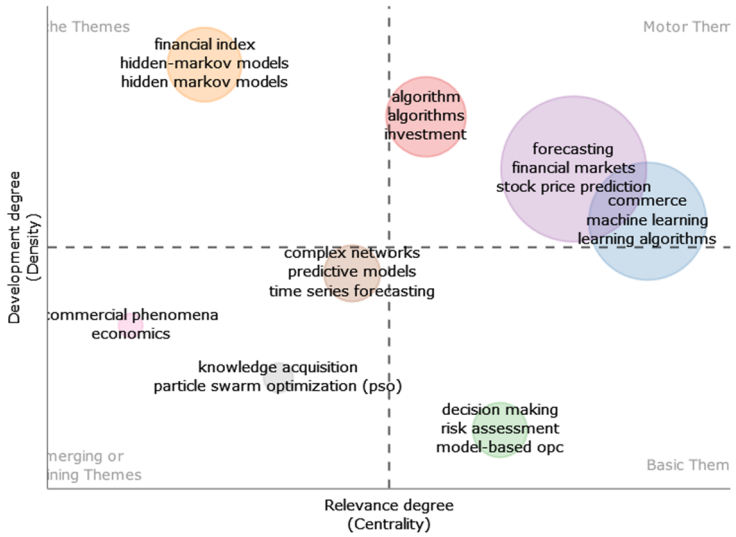

Fig.19. Thematic Map filed keywords plus

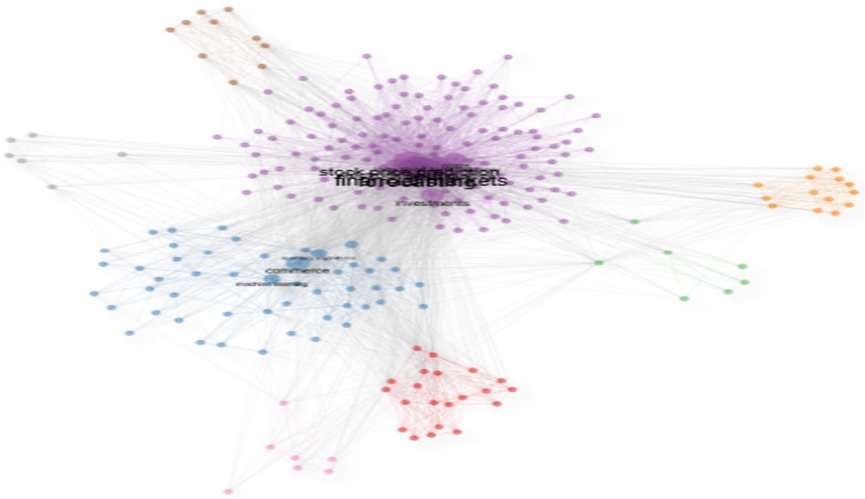

Fig.20. Thematic Map Network

1.9.1 Thematic Evolution

A "thematic map" is a kind of map that zeroes down on a single overarching concept or topic as its primary point of interest. How a specific topic or distribution differs around the world is highlighted by a thematic map. Bibliometrics use the theme map to sketch out the conceptual structure of a topic. The latter uses a co-occurrence network of phrases to evaluate the dominant topics, themes, and trends in academic discourse in a specific discipline

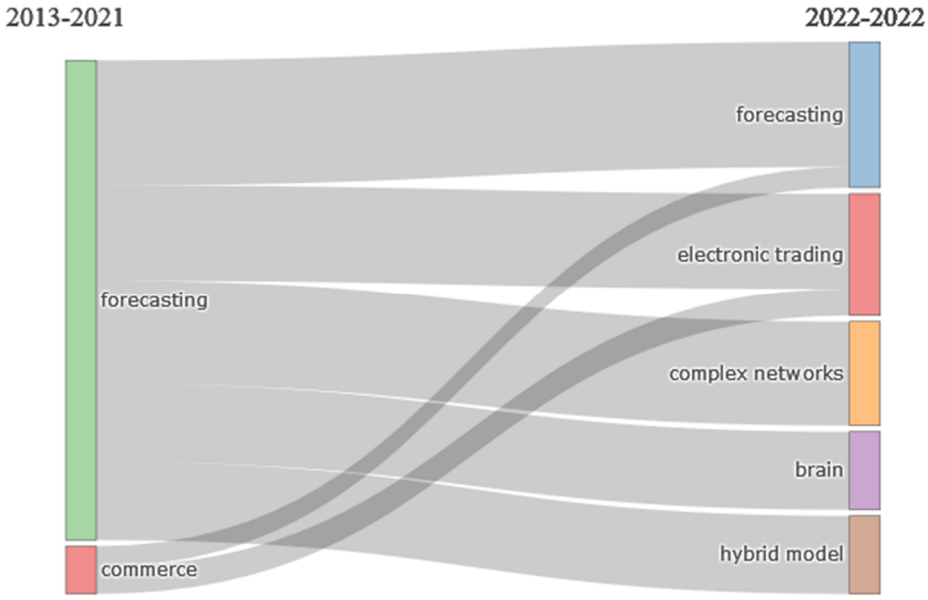

Fig.21. Thematic Evolution

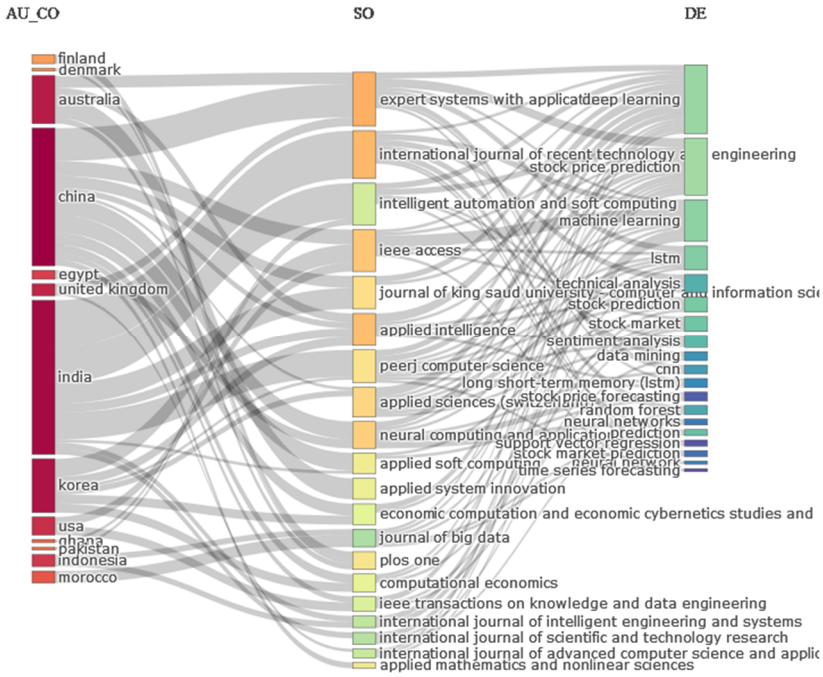

Fig.22. This presents the connectedness between the journal, the country, and keywords. It was created to depict the proportion of the participation of each country with a specific journal and topic represented by the keyword.

2 Three Map Analysis keywords Plus

Figure 23 shows the tree map used in this study by Keyword Plus. The investigation outcome demonstrates that the articles' writers employed keywords to express the magnitude of the term in the publications. The terms that are much larger than the others indicate the frequency with which the writers of the abstract use that particular word. Based on Figure 23, we can conclude that the word "forecasting" has the most significant size in this investigation, as it was used 76 times. This indicates that the writers used it as the primary phrase for forecasting throughout their work.

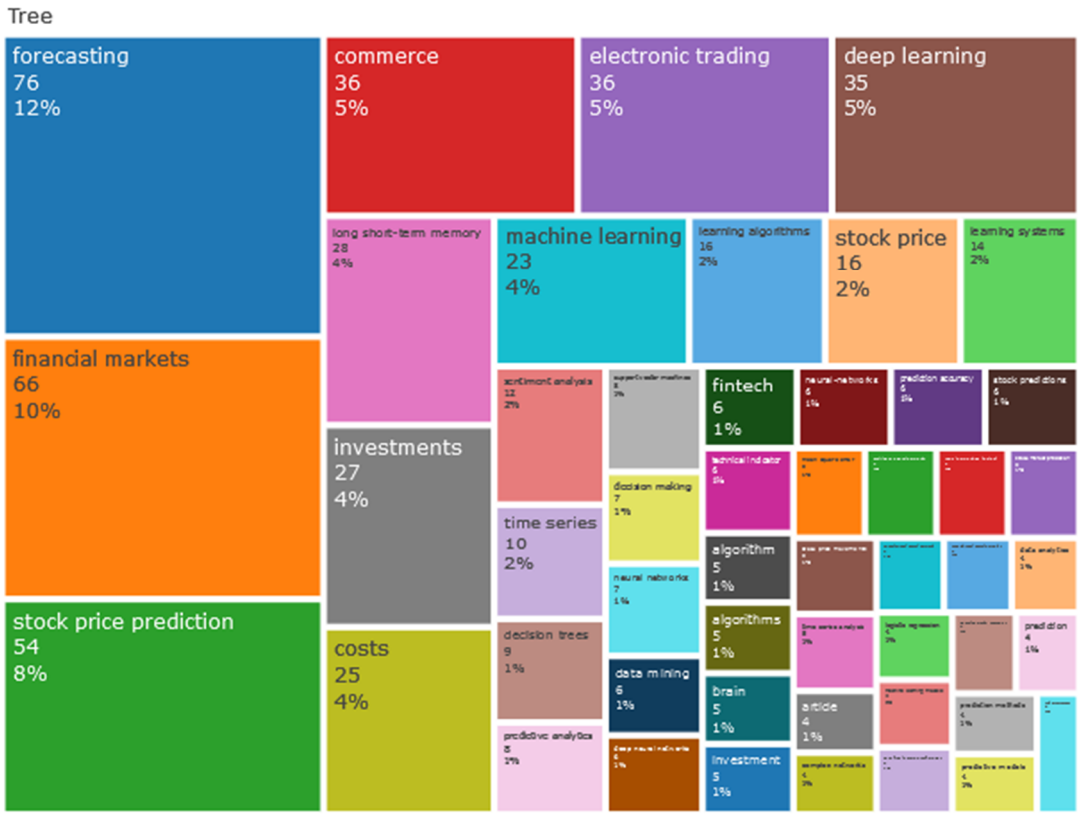

Fig.23.Tree Map Analysis

3 Factorial Approach

3.1 conceptual structure map method: MCA multiple corresponding analysis filed keyword plus

Multiple correlation analysis (MCA), shown in Figure 24, is used to assess the conceptual structure map in this research. Thus, the authors of this research have discovered a statistically significant split between the two groups. The red cluster highlights the importance of the term "stock price movement" inside the cluster and the focus of the study. Other factors, such as those related to stock price prediction and data mining, are also emphasized. The red cluster result indicates that community involvement is crucial in sentiment analysis and stock price movement.Additionally, the blue cluster represents the second grouping. The term "preferences" in the investment context is the most significant concept in the blue cluster.

These examples of algorithms—conservation and protected areas, for example—should be given more priority by the stock price prediction. For this reason, the blue cluster also stresses the park's role as the focus of investment.[14]

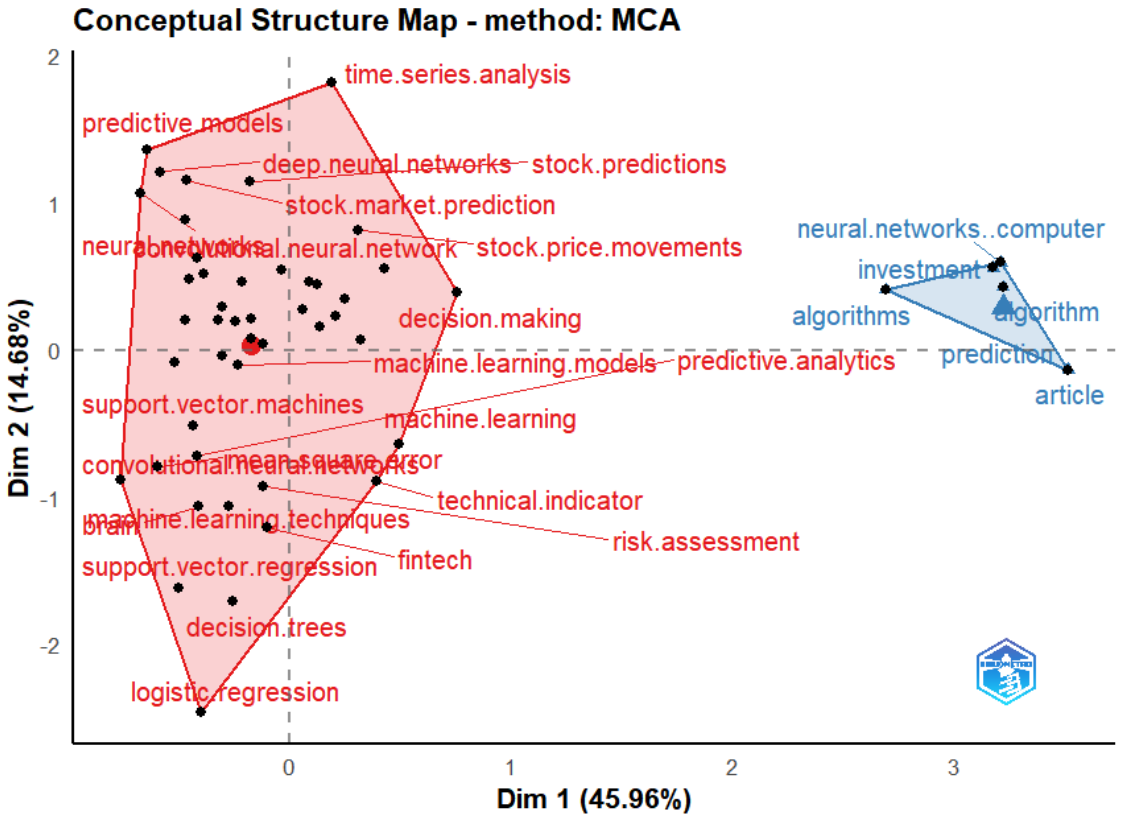

Fig.24.Multi Dimensional Scaling Analysis

3.2 Factorial Analysis topic dendrogram Multiple Correspondence analysis filed keyword plus

A dendrogram graphically represents the findings of hierarchical cluster analysis. Hierarchical cluster analysis This graphic looks like a tree, and it shows how each level of hierarchical clustering is represented as the combination of two branches of the tree into a single one. The branches are symbolic representations of the clusters produced at each stage of the hierarchical clustering process. The method of Multiple Correspondence analysis and keyword plus words by cluster include forecasting, Financial Markets, stock price prediction, commerce electronic trading, deep learning, long short-term memory, investments, costs, and machine learning.

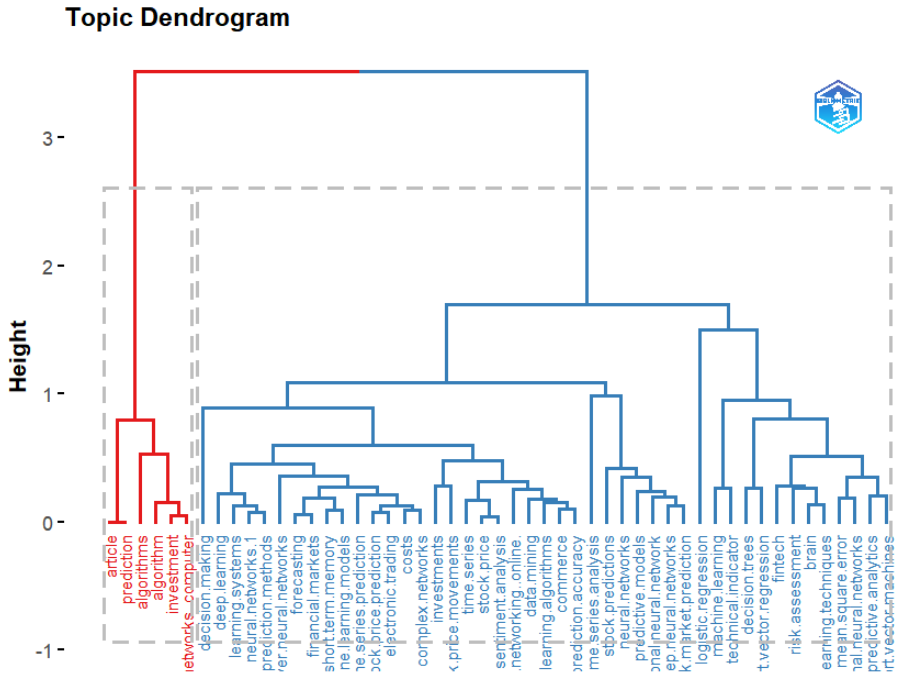

Fig.25. Factorial Analysis topic dendrogram

4. Conclusion

In conclusion, this paper provides a comprehensive and insightful exploration of the landscape surrounding Machine Learning and Deep Learning applications in Stock Price Prediction. Through an extensive bibliometric analysis spanning 2013 to 2022, this research has shed light on critical trends, influential contributors, and evolving themes within the field. The study's findings underscore the continued significance of predicting stock market movements, given its pivotal role in economic activities. Adopting advanced machine learning techniques reflects the persistent drive to enhance predictive accuracy and capture intricate market dynamics. By leveraging tools like Support Vector Machines, Neural Networks, and Reinforcement Learning, researchers have navigated the challenges posed by data complexity and market uncertainty. Analyzing publication patterns, author contributions, and country affiliations reveals an expanding landscape of research productivity. China's substantial presence in publications and citations underscores its dominance as a global contributor to this domain. This trend showcases the growing interest and commitment of countries like India and South Korea to establish their positions in this vital research area. Exploring co-occurrence and co-citation networks has highlighted influential papers and unveiled thematic shifts over the years. The emergence of "neural networks," "financial markets," and "stock price movements" as persistent and evolving themes reflects the community's focus on understanding complex market behaviors and enhancing prediction methodologies. Ultimately, this study enriches our understanding of the

intricate relationship between Machine Learning, Deep Learning, and Stock Price Prediction. By offering insights into research trends, influential authors, and emerging themes, this paper is a valuable resource for researchers, practitioners, and policy-makers. As the field evolves, this comprehensive analysis provides a foundational reference point for further exploration and advancement in the quest for more accurate stock price prediction models.

Author contributions: All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Scopus

Conflicts of Interest: The authors declare no conflict of interest.Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

[1].Huang, W., Y. Nakamori, and S.-Y. Wang, *Forecasting stock market movement direction with support vector machine.* Computers & Operations Research, 2005. **32**(10): p. 2513-2522.

[2].Moody, J. and M. Saffell, *Learning to trade via direct reinforcement.* IEEE Trans Neural Netw, 2001. **12**(4): p. 875-89.

[3].Abu-Mostafa, Y.S. and A.F. Atiya, *Introduction to financial forecasting.* Applied Intelligence, 1996. **6**(3): p. 205-213.

[4].Blank, S.C., *"Chaos" in futures markets? A nonlinear dynamical analysis.* Journal of Futures Markets, 1991. **11**(6): p. 711-728.

[5].Decoster, G.P., W.C. Labys, and D.W. Mitchell, *Evidence of chaos in commodity futures prices.* Journal of Futures Markets, 1992. **12**(3): p. 291-305.

[6].Frank, M. and T. Stengos, *Measuring the Strangeness of Gold and Silver Rates of Return.* Review of Economic Studies, 1989. **56**(4): p. 553-567.

[7].Hsu, M.-W., et al., *Bridging the divide in financial market forecasting: machine learners vs. financial economists.* Expert Systems with Applications, 2016. **61**: p. 215-234.

[8].Khalil, G.M. and C.A. Gotway Crawford, *A Bibliometric Analysis of U.S.-Based Research on the Behavioral Risk Factor Surveillance System.* American Journal of Preventive Medicine, 2015. **48**(1): p. 50-57.

[9].Yin, M., et al., *A Bibliometric Analysis and Visualization of Current Research Trends in the Treatment of Cervical Spondylotic Myelopathy.* Global Spine Journal, 2020. **11**(6): p. 988-998.

[10]. O'Sullivan, K.E., J.C. Kelly, and J.P. Hurley, *The 100 most cited publications in cardiac surgery: a bibliometric analysis.* Irish Journal of Medical Science (1971 -), 2015. **184**(1): p. 91-99.

[11]. Hagenau, M., M. Liebmann, and D. Neumann, *Automated news reading: Stock price prediction based on financial news using context-capturing features.* Decision Support Systems, 2013. **55**(3): p. 685-697.

[12]. Potter, J., *Mapping the literature of occupational therapy: an update.* J Med Libr Assoc, 2010. **98**(3): p. 235-42.

[13]. Lotka, A.J., *The frequency distribution of scientific productivity.* Journal of the Washington Academy of Sciences, 1926. **16**(12): p. 317-323.

[14]. Long, W., Z. Lu, and L. Cui, *Deep learning-based feature engineering for stock price movement prediction.* Knowledge-Based Systems, 2019. **164**: p. 163-173.

[15]. Hochreiter, S. and J. Schmidhuber, *Long Short-Term Memory.* Neural Computation, 1997. **9**(8): p. 1735-1780.