Scientometric review and analysis of recent approaches to stock market forecasting: Two decades survey

KEHINDE T.O.[1], Felix T.S. CHAN[2]*, S.H. CHUNG[1]

[1]*Department of Industrial and Systems Engineering,*
The Hong Kong Polytechnic University,
Hung Hum, Hong Kong

[2]*Department of Decision Sciences,*
Macau University of Science and Technology,
Avenida Wai Long, Taipa, Macao.

*Corresponding author: tschan@must.edu.mo (Felix T.S. CHAN)

KEHINDE T.O.: temitope.kehinde@connect.polyu.hk

Nick S.H. CHUNG: nick.sh.chung@polyu.edu.hk

Abstract

Stock Market Forecasting (SMF) has become a spotlighted area and is receiving increasing attention due to the potential that investment returns can generate profound wealth. In the past, researchers have made significant efforts to forecast the stock market trends and predict the best time to buy, sell, or hold. The essence of past investigators' various techniques and methods was to maximise the abundant opportunities that abound in the stock market trading and amass huge wealth from it. Over the years, no scientometric review has been conducted to scientifically map out the trends, progress, and limitations in the subject area. In this regard, this paper presents a pioneering scientometric review in SMF. It investigates a total of 220 reputable articles (2001-2021) to identify trends and patterns in stock market forecasting studies. VOSviewer software was used to conduct science mapping analysis. Actionable insights from the analysis explain significant metrics such as the top research outlets, most-cited articles, most co-occurred keywords, most influential countries, and much more. More so, a key finding in this paper is the introduction of a less computational approach that has the possibility of making a better forecast. Yet, past researchers have not thoroughly explored this option. This paper is beneficial to Early Stage Researchers (ESR), governments, funding bodies, managers, analysts, financial enthusiasts, practitioners, and investors, so as to understand the current progress and focus areas in stock market prediction.

Keywords: Scientometric Review; Stock market forecasting; Stock market; Stock market index; Neural network

1. Introduction

Accurate stock market models can equip investors with the information they need to make conscious decisions. Trading with these models might help anyone choose the safest, most lucrative investments. In addition, building complex models permits incorporating data like stock prices and news. Articles have been written reviewing various aspects of the financial world, such as stock market forecasting, foreign exchange forecasting, currency exchange prediction, and optimal portfolio selection. Stock price forecasting is considered difficult due to its complexity, chaoticness, and data fusion to guide future investment decisions (J. Zhang et al., 2018; L. Zhang et al., 2018). Despite the recent success of machine learning and deep learning algorithms in producing reliable stock price forecasts, there are still a number of obstacles to overcome, such as the complexity of their computations and the need to convert their non-linear models to linear ones before they can be used effectively.

The stock market, sometimes referred to as the equity market, is a pool of financial organizations where regular buying and selling of shares and securities occur. However, the volatility of such an investment makes professionals classify it as high-risk. In 2020, the value of global equity was estimated at 105.8 (USD) trillion (Katie et al., 2021). Due to the enormous value of the global stock market, investors are always on the lookout for opportunities to profit from financial markets as long as corporations keep listing their shares. In the past, experienced investors have relied on logical discernment to characterise market trends and patterns; however, this method is basically archaic. Due to the evolution of a large amount of data associated with global stocks, the old method of judging based on discernment that lacks a theoretical proof can no longer work. More so, analysts have used simple statistical techniques to gain insights and deduce market trends. Today, more sophisticated methods have evolved, e.g., Artificial Neural Network (ANN) (Omidi et al., 2011), which is the most used technique in the modern era of stock market forecasting (SMF).

The innovations in SMF have earned the curiosity of many shareholders and analysts. The stock market system is complicated due to its dynamic trending nature, volatility, and characteristic noisy environment (Araújo & Ferreira, 2013). Several economic and non-economic factors affect the market trends, the varying factors that influence stock prices include trader expectation, future income, newscast on returns, financial circumstances, pronouncement of dividends, inflation, recession, administrative events, change in management, quarterly earnings reports, news, and many more. Despite using these indicators to predict stock returns, making an accurate daily or weekly forecast of the stock market is still a challenge.

The accurate forecasting of the stock market is a thrilling and complex task in the fluctuating market domain. A few decades ago, analysts used traditional techniques to predict prices based on historical data. However, this approach is less effective due to some particular features that financial markets possess, such as noise, volatility, anomalies, and fluctuating trends. Time series models such as Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), Generalised ARMA (GARMA), and Auto-Regressive Conditional Heteroscedasticity (ARCH), and Generalised ARCH have become popular among analysts and researchers (Bollerslev, 1986; Cheng et al., 2010; Hyndman & Athanasopoulos, 2018). Nonetheless, due to the limitation of these models in handling the chaotic and volatile nature of stock market system, researchers have advanced to using conventional time series models, e.g. fuzzy sets. In addition, many intelligent techniques such as Genetic Algorithms (G.A.), Neural Networks (N.N.), Back Propagation Algorithms (BPA) have been helpful and valuable to researchers in predicting the financial market (Gandhmal & Kumar, 2019).

Kimoto et al. (1990) developed a model for forecasting stock market trends using the time series model. Nikolopoulos and Fellrath (1994) combined G.A. and Neural Network to design a hybrid model to make proper investment decisions. Similarly, in Kim and Han (2000), G.A. was used with ANN to effectively forecast stock market index. In another context, Neural Networks and Time Series Model were hybridised in Roh (2007) to forecast the stock value direction and deviation variability. Using decisions based on profit and loss, soft computing is another area that has gained ground in making an accurate and effective prediction of stock market indices. In Ibidapo et al. (2017), the authors provide a coherent overview of the application of soft computing in SMF. Several researchers have used Fuzzy Logic (F.L.), Particle Swarm Optimisation (PSO), ANN, and Support Vector Machine (SVM) to accurately forecast and analyse market trends and stock prices (Ghiassi et al., 2013; Hassan, 2009; Huang et al., 2005; Kara et al., 2011; Majhi et al., 2008; Sheta et al., 2015; Thavaneswaran et al., 2009; Zhou et al., 2018). Today, the stock market has become a capital market where advanced tools are needed to monitor, predict, and regulate the market. Stock market prediction solely depends on data from the investments and trades done in the market. Advanced tools are needed to make an accurate prediction that can enable correct decision making.

Despite the popularity of machine learning algorithms in recent years, utilizing them to make predictions about several stocks might be difficult. Using machine learning algorithms to forecast multiple stocks can be challenging. Additionally, machine learning fails to distinguish between performing and non-performing stocks easily. A well-known method for dealing with this problem is Data Envelopment Analysis (DEA), which is helpful if several inputs and outputs features of multiple stocks are considered. DEA allows for a comprehensive, comparative analysis of numerous equities. Since conventional DEA models are not equipped to deal with stock forecasting, there is a need to consider an extended form of classical DEA that can fill this gap. Inverse DEA has been developed through an extension or modification of the standard DEA. Inverse DEA models are helpful in solving innovative inverse problems, such as the prediction of stock return.

The growing concern of any investor is to build an efficient and effective model to forecast what the future market holds. The previous related works explore various means of analysing the stock market. To our best knowledge, a scientometric review has not been conducted in this research area. A scientometric analysis is an appropriate fair approach for achieving a holistic assessment to unravel previous contributions, progress, trends, and limitations in a chosen area of research. A scientometric analysis is a method that analyses trends and progress in a particular field of study quantitatively and critically. It gives the literary mapping of the existing structure of related articles by quantitatively analysing journals, books, conference papers, and other publications (Blažun et al., 2015).

Several studies adopt different techniques when reviewing the literature; for example, techniques such as Comprehensive Review (Cavalcante et al., 2016; Kurani et al., 2021; Thakkar & Chaudhari, 2021a), Quantitative Analysis (Sharma & Kaushik, 2018), Empirical Analysis (Li et al., 2016), and Systematic Review (Gandhmal & Kumar, 2019; Jabbar Alkubaisi et al., 2017; Li & Bastos, 2020; Nti et al., 2020; Thakkar & Chaudhari, 2021b) have been adopted within the context of SMF. Despite their in-depth reviews, most are manuals and often are not reproducible. Besides, according to Markoulli et al. (2017), manual studies broadly explore the trees without considering the forest.

In this study, to achieve a wider picture of bibliometric data, tools and methods, scientometric research has been conducted to review the progress and trends in SMF, to complement and add to the existing knowledge on stock market prediction. Utilising one of the best academic databases, i.e. Scopus, we retrieved all the related academic articles published over 20 years. Scopus is a well recognised academic database enriched with data that link scholarly articles across all disciplines. The selected databank retrieves data for mapping and analysing the network of various occurrences and collaborations in SMF. The reason for this mapping is to systematically and critically scrutinise previous studies and identify the impacts made so far in the years under coverage. This will enable us to spot the targeted areas in the earlier studies.

Furthermore, this study will inform practitioners and researchers of the latest advancements in SMF. This article combines Sciencetometric analysis with an insightful discussion on the tactical approach so as to have a holistic perspective and gain more groundbreaking SMF results. Most importantly, academia and industry will find this work helpful in exploring more hybridised or advanced techniques to predict the stock market better.

2. Research methods

Scientometric analysis provides a visual perspective of the trends and progress in a chosen area of study. It spots the collaborators in a field of interest, thereby giving room for a complete understanding of the research gap to explore. This section highlighted the tools and methods used in carrying out the mapping analysis. The flowchart in Fig. 1 depicts the general overview and brief outline of elements discussed.

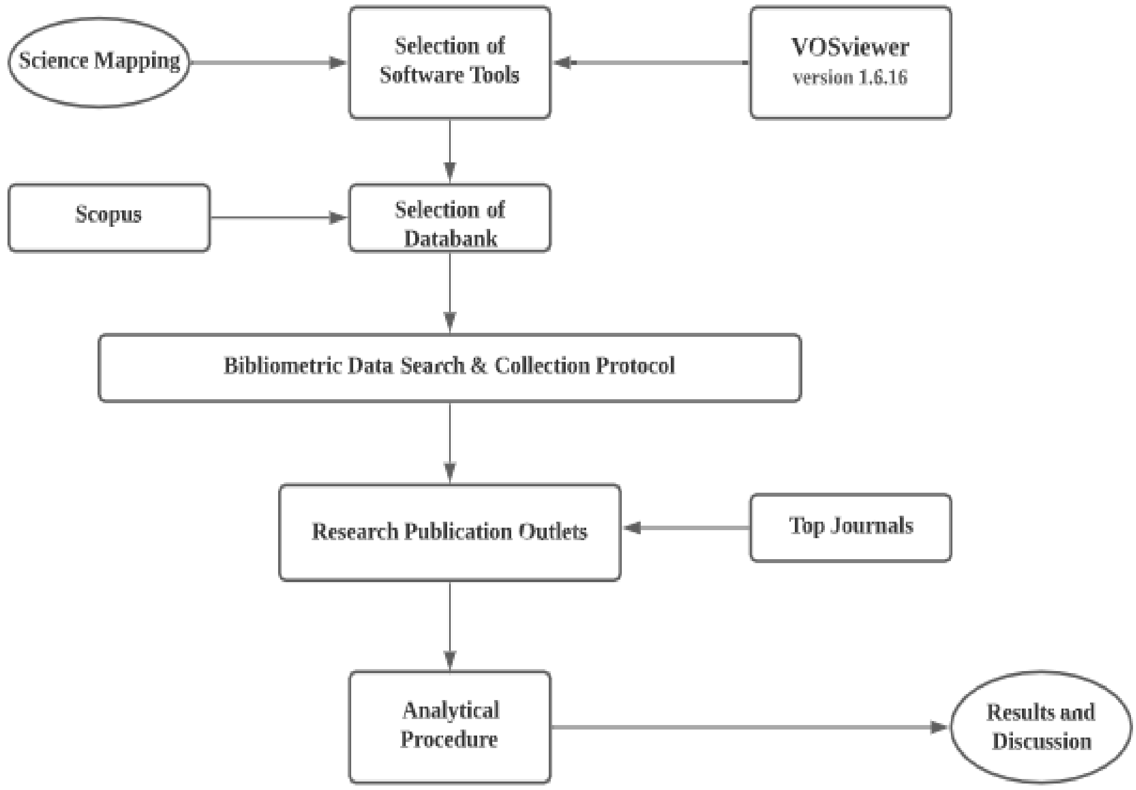

Fig. 1: Research methodology flowchart.

2.1. Science mapping and scientometric techniques

Science mapping is a technique employed in bibliometric data analysis and network visualisation of scholarly knowledge and contributions in a chosen area of study. It explores a body of scientific literature, statistical tools, metrics and indicators to highlight significant trends and patterns for future discourse (Chen, 2017). The purpose of using science maps is to showcase how publications, authors, affiliations, and keywords relate to one another (Osinska & Malak, 2016; Rafols et al., 2010; Van Raan, 2019), so that key research areas and key actors in a chosen subject area are easily identified. Science mapping is a tool used to identify, describe, visualise, and analyse past research articles for specific purposes (Tijssen & Van Raan, 1994). Notably, this technique helps to lessen the stresses and challenges involved in carrying out manual reviews, which is often tedious, time and energy-consuming while trying to map out the connections between keywords, authors, outlets, countries, and so on within a chosen domain of research interest (Su & Lee, 2010).

As a bottom-up approach system, Science mapping analysis answers questions such as "What are the key areas within a specified research domain?", "How do these areas relate to one another?” “What is the distribution of publications over the years?” “Who are the key players (in terms of authors, affiliations and journals)?” The answers to these questions are invariably relevant for science policy purposes. In addition, Science mapping helps draw out mutual association and links thre strength among keywords, authors, affiliations, and countries in any particular scientific domain. Generally, science mapping helps in knowledge organisation (K.O.). It is categorised into three types, namely informatics, bibliometrics, and scientometrics; however, these techniques are often overlapping (Wuni et al., 2019). Many researchers prefer scientometric analysis to analyse bibliometric data in order to gain broader insights into the potential trends and patterns. In contrast, bibliometrics chiefly concentrates on the literature per se (Hood & Wilson, 2001).

For the purpose of this study, the third category of science mapping met the study’s objective. Even though Science mapping is seen as a part of the K.O. system, it is not meant to replace the traditional taxonomy and classificatory schemes (Hjørland, 2013; Mazzocchi, 2018). In the course of this study, and as described by Cobo et al. (2011), we employ a quantitative science mapping approach to map networks, patterns, and models from a massive set of bibliometric data.

2.2. Selection of software tools

Over the years, researchers have used various software in conducting scientometric reviews of works of literature (Cobo et al., 2011). Most of these tools are free and readily available online for various window versions. In addition, it is pertinent to know that the usage of these tools varies, as some are intended for specific purposes, others are meant for general purposes. Each tool has its strengths and weaknesses, which makes it unique. Appropriate consideration must be ensured in choosing the right tool that meets the aim of the study. In this realisation, it is essential to fully understand each tool’s features, strengths, and limitations to ensure appropriate selection. The commonly used science mapping tools include IN-SPIRE, BibExcel, BiblioMaps, CiteSpace, Gephi, Biblioshiny, VOSViewer, Bibliometrix, CoPalRed, VantagePoint, CitNetExplorer, SciMAT, Leydesdorff’s Software, Sci^2 Tool, HistCite, and Network Workbench tool (Bastian et al., 2009; Cobo et al., 2011, 2012; Garfield, 2004; Liu et al., 2013; Moral Muñoz et al., 2020; Van Eck & Waltman, 2010, 2014). Van Eck and Waltman (2010) developed the first version of VOSviewer for mapping out similarities in visualisation. In our study, VOSviewer (v1.6.16) software was chosen to systematically map and visualise SMF trends over two decades. The choice of VOSviewer is a result of its consistency in results, user-friendliness and Graphical User Interface (GUI). Besides, VOSviewer is open-source software, which has been proven to be a better alternative to the Multi-Dimensional Scaling (MDS) technique (Van Eck et al., 2010). Renowned researchers have widely used VOSviewer for various Science Mapping reviews. For example, VOS has been used to map out knowledge in notable works of prominent researchers, such as Fabregat-Aibar et al. (2019), Gaviria-Marin et al. (2018), Hosseini et al. (2018), Li and Xu (2021), Somanathan and Rama (2020), Van Eck and Waltman (2010) and Van Eck and Waltman (2017).

2.3. Selection of databank

Choosing an appropriate database and employing appropriate methods are essential to the success of any scientific review (Oteng et al., 2021). However, because their coverage for various research disciplines differs, their results may vary (Mongeon & Paul-Hus, 2016). Scopus and Web of Science are among the most trusted and credible databanks that index the Stock market and other scientific research. They both offer comprehensive repositories of various bibliometric data. In affirmation of this, previous reviews on stock market forecasts also downloaded bibliometric data from either or both sources, e.g. Bustos and Pomares-Quimbaya (2020), Ferreira et al. (2021), Hassan et al. (2020). This study selects Scopus because it has broader coverage, with the latest articles, than Web of Science (Meho & Rogers, 2008). In addition to this, Scopus is preferable to Web of Science because it archives a large number of academic documents than its counterpart (Aghaei Chadegani et al., 2013; Guz & Rushchitsky, 2009).

Scopus is chosen over a database like Google Scholar because of its reliability and the credibility that it houses non-predatory journals. More so, the idea of exploring a single source eliminates duplication and redundancy that may complicate the analysis process. VOSviewer supports direct usage of bibliometric data in CSV file format; hence, the choice of Scopus as a unified database is consistent with the corresponding format of VOSviewer (1.6.16).

2.4. Bibliometric data extraction and journals selection

In this section, prominent articles such as Atsalakis and Valavanis (2009), Cheng et al. (2007), Ferreira et al. (2021), Gandhmal and Kumar (2019), Guresen et al. (2011), Jiang (2021), Schumaker and Chen (2009), Somanathan and Rama (2020), Thakkar and Chaudhari (2021b) were used as a guide in choosing the most appropriate keywords to suit this study. In this connection, notable keywords used in the search and retrieval of bibliometric data are "stock price" OR "stock index" OR "stock market" OR "stock market prediction" OR "stock prediction" OR "stock forecasting" OR "stock market forecasting" OR "stock market indicator" OR "stock market trend," The preliminary search was done using a combination of these keywords, which sum up the query: TITLE ("stock price" OR "stock index" OR "stock market" OR "stock prediction" OR "stock market prediction" OR "stock forecasting" OR "stock market forecasting" OR "stock market trend" OR "stock market indicator"). The preliminary search query generated 18482 articles (as of 11th July 2021). Next, the search query was further refined to cover the scope of the study by deploying two more levels to the existing single search level. By doing this, the query used becomes: TITLE ("stock price" OR "stock index" OR "stock market" OR "stock prediction") AND TITLE-ABS-KEY ("stock market prediction" OR "stock forecasting" OR "stock market forecasting" OR "stock market trend" OR "stock market indicator") AND TITLE-ABS-KEY ("ANN" OR "neural network" OR "machine learning" OR "deep learning" OR "Bayesian model" OR "model fusion" OR "information fusion" OR "feature fusion"), and this gave 571 articles. Further, the document type is limited to reviews and articles to ensure robust analysis and to extract only the relevant datasets. This measure is taken to avoid complications and duplications from conference papers and books (Butler & Visser, 2006). Even though some researchers advocate the inclusion of conference proceedings in conducting reviews, this study captures journal articles only because journals are more reliable and reputable sources of knowledge (Jin et al., 2018; Zheng et al., 2016).

Lastly, the language is checked to restrict retrieved articles to English. English language is selected because it is the world's most popular and acceptable language, and most institutions and top-ranked journals only publish in this language. As a result, 220 articles were generated, including 209 research articles and 11 reviews, on 11th July 2021. The bibliometric data was received in CSV and RIS file format. The CSV file was recruited into VOSviewer for mapping analysis to gain meaningful insights into stock market prediction trends, while the RIS file was imported into EndNote for bibliography management.

2.5. Analytical Procedure

Following data importation into science mapping software, analyses were carried out at distinct stages to ensure high process fidelity. Careful review was performed to rule out the possibility of error in the interpretation of the data. Proper scrutiny was provided to avoid any discrepancy in the analysis of collected data. At first, after the VOSviewer software was launched, the data type was set at "map-based bibliographic data", then the data source was checked at “bibliographic database file”, proceeding to upload the “Scopus CSV file” which helps in carrying out analysis and mapping activities like citation, co-citation, co-authorship, co-occurrence, and bibliographic coupling. Next, each option was further examined under the “Unit of Analysis” using indicators like documents, sources, authors, institutions, and countries. The subsequent section reports and demonstrates the result and various analyses conducted on the recruited data.

3. Science mapping, results and discussions

3.1. Annual distribution of publications

Though research on the stock market has evolved since the 1930s (Somanathan & Rama, 2020), it appears that studies on SMF began two decades ago. Our study addresses 220 documents published between 2001-2021. Despite that the query not being filtered to a particular date range, it is evident from our Scopus output that the earliest publication in stock forecasting appeared in the year 2001, in the Asia-Pacific Journal of Operational Research (Phua et al., 2001), where Phua *et al.* (2001) applied the Neural Network (N.N.) algorithm in forecasting the stock market.

Fig. 2 shows that the yearly publication movement was creeping along in the 2000s until it accelerated in 2005. This was expected as pioneer research requires robust models to accurately forecast the market due to its volatility and dynamic nature. Although it appears that there were dips in 2003, 2006, 2008, 2011, 2013, 2015, 2017 and 2020, and in most cases, the descent only lasted a year. Also, it is evident that the research area skyrocketed in 2019 with an all-time high value of 42 articles. This peak value recorded in 2019 is due to the growing concerns about combining techniques to outperform existing stock market prediction models. In 2020, there was slippage, as the number of published articles recorded a decline. This decline is attributed to the unprecedented event of the Covid-19 pandemic. As of mid-year in 2021, a total number of 34 publications were recorded, which suggests that there is a great prospect of a new all-time high record before the year runs out.

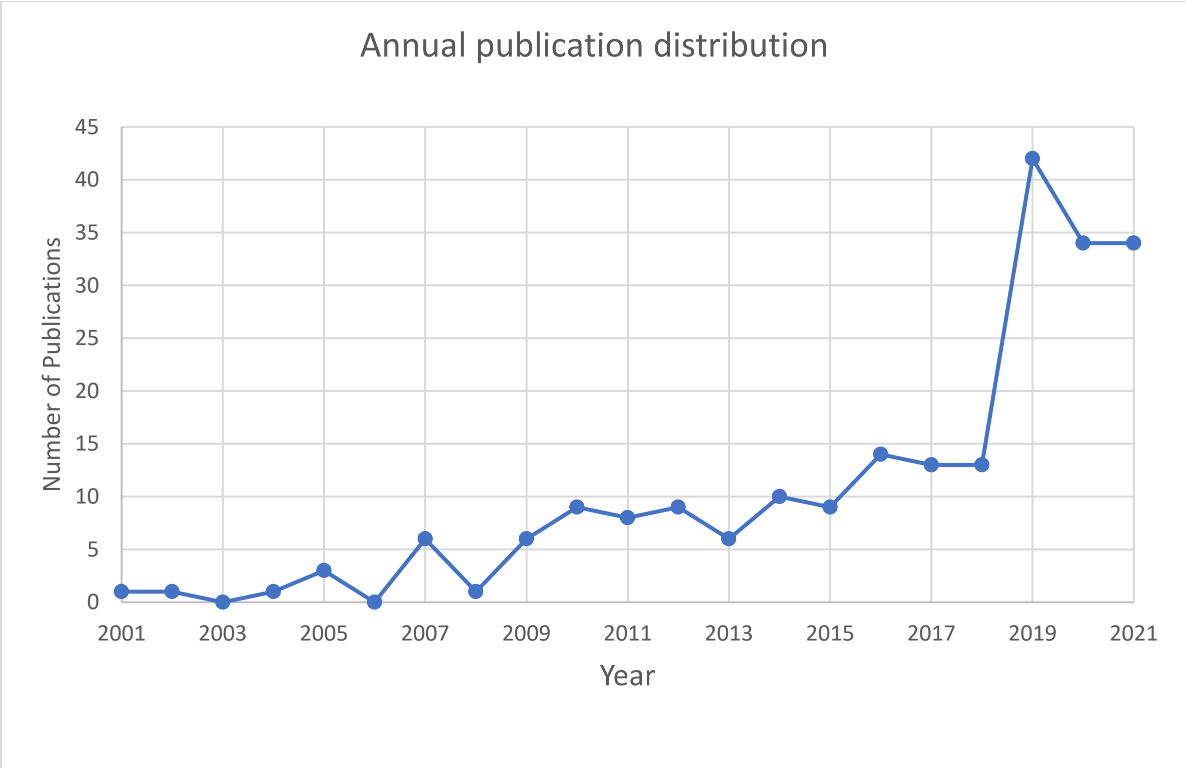

Fig. 2: Annual distribution of publications.

3.2. Top research outlets

Publishing in research outlets is a leading method to disseminate academic discoveries and breakthroughs. These outlets publish cutting-edge researches based on discipline, scope, and other notable metrics. Key journals are crucial in analysing the trends in an area of interest. In this study, VOSviewer systematically mapped out key journals by using appropriate metrics and indicators. To achieve this, the metric was set at 5 articles and 40 citations. The metrics used in mapping analysis were not based on any benchmark, but based on previous scientometric reviews, in conjunction with authors' discretion and understanding (Bello et al., 2021). These parameters generate the network map of the leading journals. Table 1 summarises the results, showing details like the number of published articles, total citations (T.C.), average citations (A.C.), total link strength (TLS) and normalised citations (N.C.).

Table 1: Leading research outlets in SMF.

Journal Name	No. of Articles	T.C.	AC	TLS	NC
Expert Systems with Applications	21	2676	127	18	62.64
Knowledge-Based Systems	5	413	83	7	11.07
Neurocomputing	5	279	56	6	7.60
Neural Computing and Applications	5	141	28	5	8.16
IEEE Access	11	84	8	11	9.27
Journal of Theoretical and Applied Information Technology	5	41	8	1	1.36

In relation to the frequency of publications, which is a means of measuring productivity, the journal of "Expert Systems with Applications" and "IEEE Access" had 21 and 11 outputs, respectively. They were the top (2) research outlets that published most articles in SMF. Table 1 shows that the remaining journals in the table only published 5 articles each, ranking them equal, based on the frequency of publications. Further, the authors assessed the research outlets to estimate their influence and impacts in the scientific world. The results of the analysis in terms of quartile ranking (Q-Index), Scimago Journal and Country Rank (SJR) and Hirsch Number (H-Index) are shown in Table 2.

Table 2: Scientific influence of research outlets (Details retrieved from scimagojr.com/journalrank.php on 6th August, 2021).

Journal Name	Q-Index	SJR	H-Index
Expert Systems with Applications	1	1.37	207
Knowledge-Based Systems	1	1.59	121
Neurocomputing	1	1.09	143
Neural Computing and Applications	1	0.71	80
IEEE Access	1	0.59	127
Journal of Theoretical and Applied Information Technology	4	0.15	29

From Table 2, out of the 6 outlets recorded, 5 belongs to the best quartile, i.e. Q1, while 1 belongs to the fourth quartile, i.e. Q4. Also, judging by the H-Index, a metric that measures productivity and impact of scholarly articles, the “Expssert Systems With Applications” journal is seen to be the best, with Hirsch number 207 (as of 6th August 2021).

Similarly, using the source citation network of the VOSviewer, Fig. 3 was generated. The node size denotes the number of citations, which measure the impact made by the leading outlets in the chosen area of study. The bigger the node, the higher the impact.

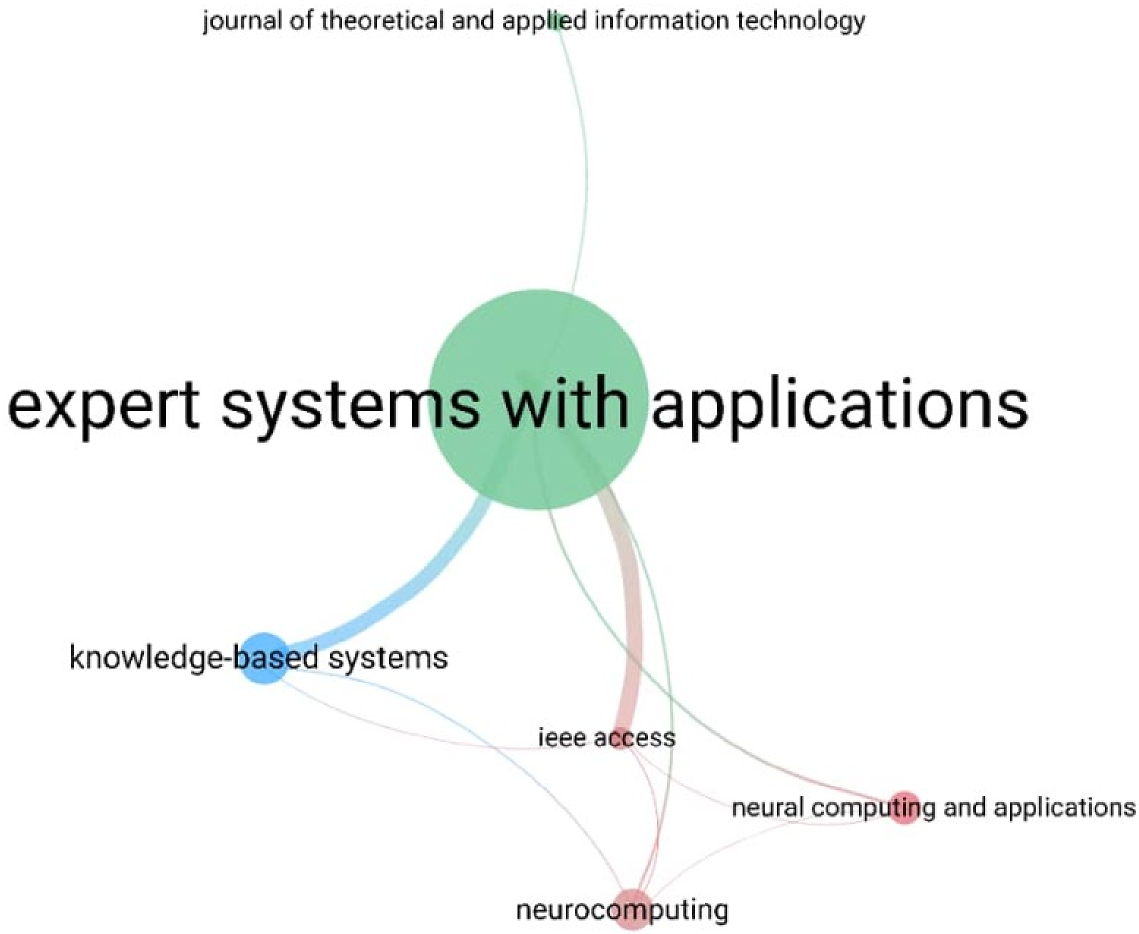

Fig. 3: Network of top research outlets.

Fig. 3 shows that the node size of “Expert Systems with Applications” is relatively larger than all the other nodes, indicating that “Expert Systems with Applications” made the most significant impact in the area of SMF, with a total citation of 2676. After “Expert Systems with Applications”, the second, third, fourth, fifth and sixth ranking of citation impacts are “Knowledge-Based Systems”, “Neurocomputing”, “Neural Computing and Applications”, "IEEE Access", and "Journal of Theoretical and Applied Information Technology" with 413, 279, 141, 84, and 41 citations respectively.

Similar outlets share the same colour and belong to the same cluster; meanwhile, the lines imply the strength of connections based on cross-citation. For instance, in the red cluster, "Neurocomputing" shows stronger citation strength with "Neural Computing and Applications" and "IEEE Access". This indicates that these outlets are mostly cross-cited in many scenarios. Also, the reason for the similar clusters may be that similar research scopes guide the outlets. For the green cluster, the journals in this category include "Expert Systems with Applications" and "Journal of Theoretical and Applied Information Technology". Lastly, we have a sole outlet in the blue cluster, which is "Knowledge-Based Systems."

3.3. Co-occurrence network of keywords

Keywords, reflecting the theme of research publications, are crucial tools for retrieving appropriate journals. Just as it was done during the bibliometric data extraction, selecting the right keywords could facilitate the accomplishment of various research goals. This section analyses the co-occurrence network of authors' keywords from the retrieved 220 articles. The recommendations by Hosseini et al. (2018), Jin et al. (2018), Glänzel and Schubert (2004), and Wang et al. (2019) were adopted to accomplish this aim. The analysis was carried out by checking the co-occurrence function and setting the unit as "Authors' keywords", and then the counting option was checked at fractional counting. The default number of keywords was chosen, i.e. 5. Out of 635 keywords recorded from the 220 downloaded bibliometric data, only 33 met this threshold. The detailed analysis of the results is duly presented in Table 3. Table 3, with 33 of the most commonly used authors' keywords, occurrences (O) and total link strength (TLS).

Table 3: Commonly used authors' keywords.

S/N	Keywords	0	TLS	S/N	Keywords	0	TLS
1	Stock Market Prediction	50	58	**18**	Feature Selection	8	16
2	Stock Market	36	68	**19**	Genetic Algorithms	8	10
3	Machine Learning	35	67	**20**	Artificial Neural Networks	8	9
4	Deep Learning	28	48	**21**	Support Vector Machine	7	16
5	Stock Market Forecasting	19	18	**22**	Technical Indicators	7	14
6	Neural Network	17	29	**23**	Artificial Intelligence	7	10
7	Artificial Neural Network	17	27	**24**	Stock Markets	6	13
8	Genetic Algorithm	16	36	**25**	Stock Price Forecasting	6	5
9	Sentiment Analysis	15	29	**26**	Finance	5	16
10	Prediction	14	27	**27**	Natural Language Processing	5	13
11	Data Mining	13	25	**28**	Fuzzy Logic	5	12
12	Forecasting	11	21	**29**	Time Series	5	10
13	Neural Networks	11	20	**30**	Time Series Forecasting	5	10
14	Stock Price Prediction	11	13	**31**	Text Mining	5	8
15	Stock Prediction	10	15	**32**	Support Vector Regression	5	7
16	Classification	9	16	**33**	Stock Forecasting	5	4
17	LSTM	8	20				

From Table 3, the most used keyword was "stock market prediction", with a total occurrence of 50, followed by "stock market" (36) and "machine learning" (35). Understanding these keywords can guide researchers on the proper index, retrieval, and usage of choice of words. Calculating the correlation coefficient between O and TLS, a strong positive correlation ($r = 0.927$) exists between the two indices. Fig. 4 presents the mapping of all the 33 authors' keywords extracted. The keywords that are closely associated with one another are packed into distinct clusters. The size of each node represents the frequency of occurrence in which several authors have used a particular keyword.

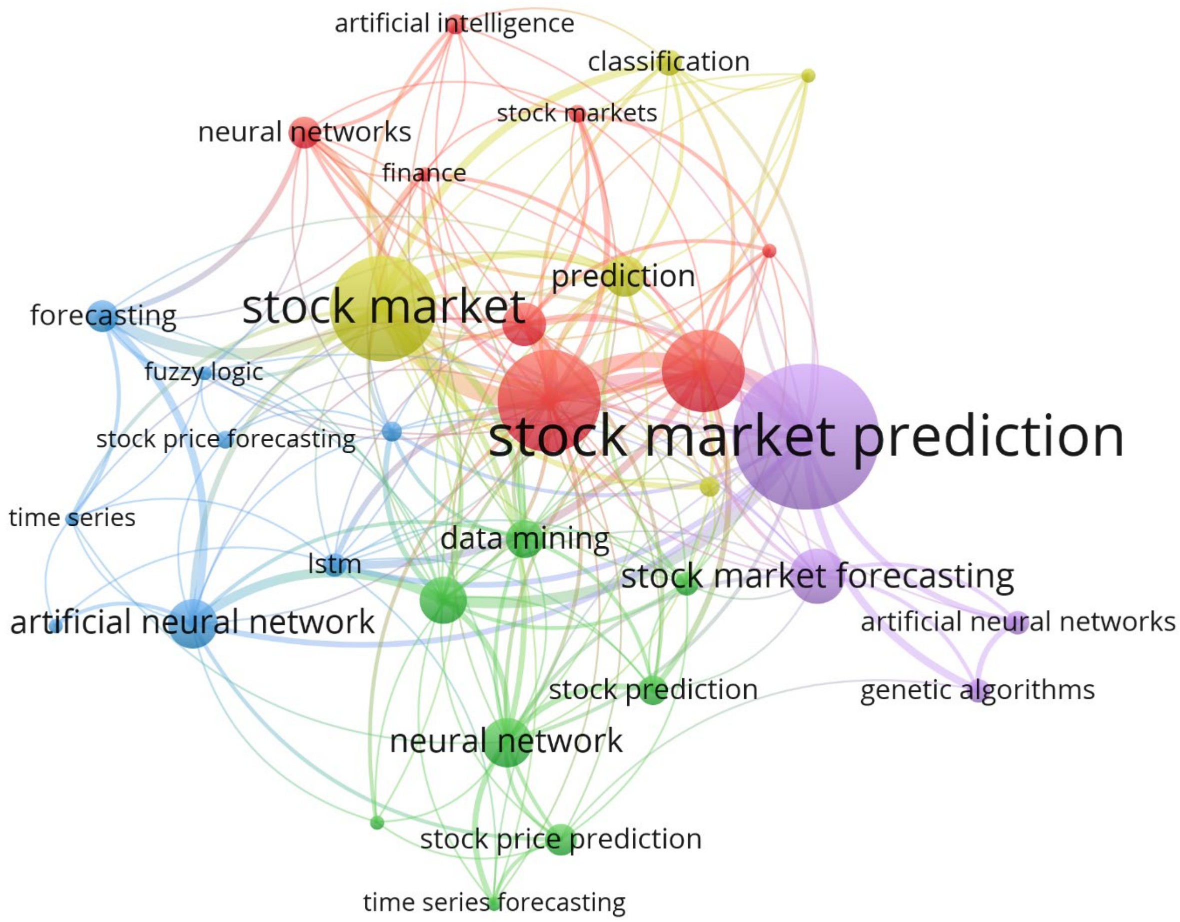

Fig. 4: Authors' keywords analysis.

Cluster Analysis

Red Cluster: the keywords in this category are "machine learning (ML)", "deep learning (DL)", "natural language processing (NLP)", "sentiment analysis", "finance", "stock markets", "neural networks", and "artificial intelligence(A.I.)."

Blue Cluster: the keywords in this category are "artificial neural network (ANN)", " Long Short Term Memory (LSTM)", "support vector machine (SVM)", "fuzzy logic (F.L.)", "forecasting", "stock forecasting", "stock price forecasting", and "time series."

Green Cluster: the keywords in this category are "stock prediction", "stock price prediction", "feature selection", "data mining", "genetic algorithm (G.A.)", "time series forecasting", "support vector regression (SVR)", and "neural network.

Yellow cluster: the keywords in this category are "stock market", "technical indicators", "classification", "text mining", and "prediction."

Purple cluster: the keywords in this category are "stock market prediction", "stock market forecasting", "genetic algorithms (G.A.)", and "artificial neural networks (ANN)."

The overview of all the clusters is that most keywords illustrate how the stock market is predicted using various machine learning and deep learning techniques, such as ANN, SVM, GA, LSTM, FL, sentiment analysis, technical analysis, and many others.

3.4. Co-authorship network analysis

In academia, articles published without collaboration is tantamount to lower productivity (Glänzel & Schubert, 2004). Thus, the partnership among investigators and educational establishments enriches valuable ideas, which habitually leads to significant landmarks and productivity in research outputs. This section discusses the co-authored research articles to determine the leading researchers in the area of interest. Generally, most articles are published with multiple authors; this analysis reveals the vital collaborations among several authors in SMF research. The metric used in this analysis is a minimum of 3 articles with at least 15 citations. Intriguingly, no more than 13 authors met this benchmark out of 555 authors. Fig. 5 illustrates the overlay visualisation network of co-authorship among leading researchers in SMF.

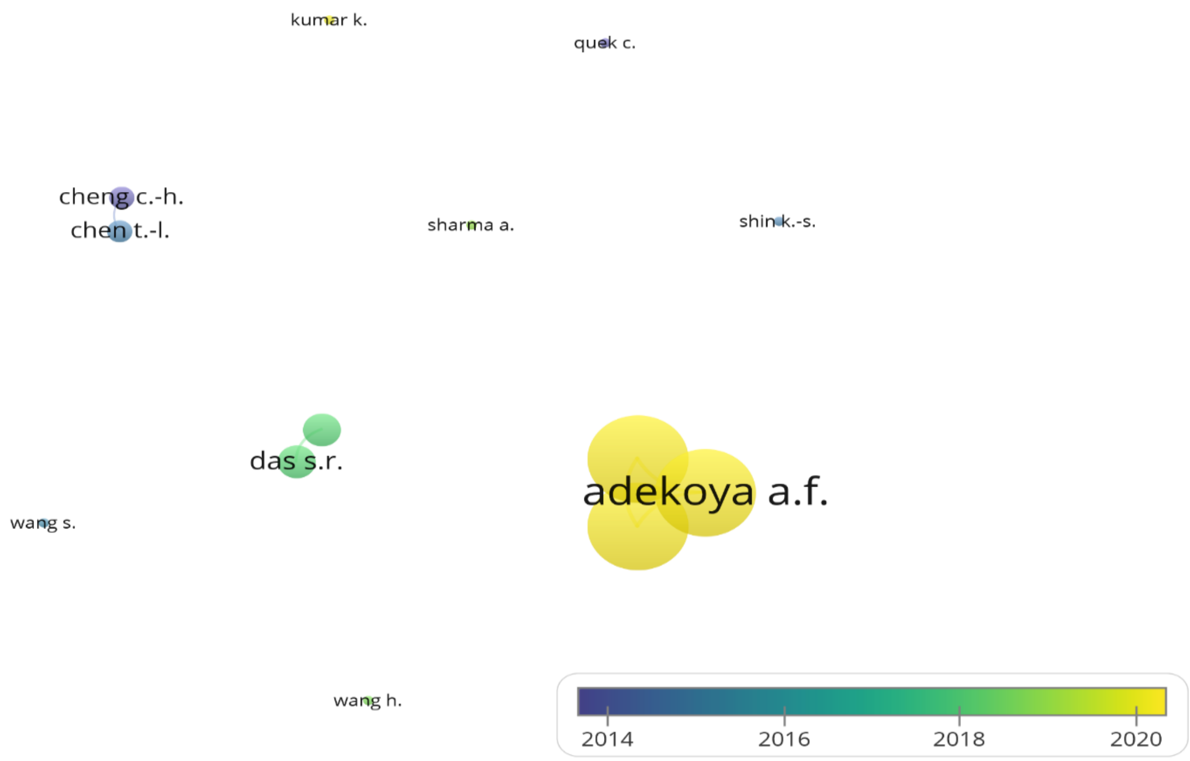

Fig. 5: Collaboration network of authors in SMF.

As seen in Fig. 5, the most collaborative strength was among three researchers, namely Adekoya, Weyori, Nti. One of the notable works where the top three authors collaborate is in Nti et al. (2020), where the authors conducted a systematic-critical review of articles in stock forecasting. More so, it is evident that the collaborative strength between these three authors occurred recently, less than five years ago. Their findings further support the claim that ANN and SVM are the most explored algorithms in stock forecasting. The total link strength is used to filter the command prompt. Table 4 further buttresses the analysis regarding the number of articles published, T.C., A.C., and TLS.

Table 4: Most collaborative authors analysis.

Author	No. of Articles	TC	AC	TLS
Adekoya A.F.	4	29	7	8
Nti I.K.	4	29	7	8
Weyori B.A.	4	29	7	8
Das S.R.	3	15	5	3
Mishra D.	3	15	5	3
Chen T. L.	3	54	18	2
Cheng C. H.	3	29	10	2
Kumar K.	4	16	4	0
Sharma A.	4	17	4	0
Wang H.	4	44	11	0
Quek C.	3	47	16	0
Shin K. S.	3	202	67	0
Wang S.	3	216	72	0

3.5. Network analysis of article citation

Over the years, bibliometric performance has been viewed as a measure of citation counts (Wang et al., 2019; Zhang, 2013), and as a result, most publications were usually assessed by their frequency of citations. In academia, scholarly reputation is measured chiefly by the impacts of research, often determined by the sum of citations of an article. Even though this notion is subject to dispute, research articles with high citations are often more valued and credited as impacts to the authors of such documents. Citations give impact, while the frequency of publications gives productivity. Invariably, a researcher can choose to be impactful rather than productive and vice versa. To analyse the articles that have made significant impact in SMF, the threshold of analysis is set intuitively at 40 citations. When it comes to impact, quality speaks and not quantity, so our attention is to emphasise the citation counts, while less or no emphasis is placed on the frequency of publications. By so doing, the specified threshold generated 34 documents out of 220 analysed articles. However, out of 34 documents, only 24 were connected. Fig. 6 demonstrates the density map of the connected articles.

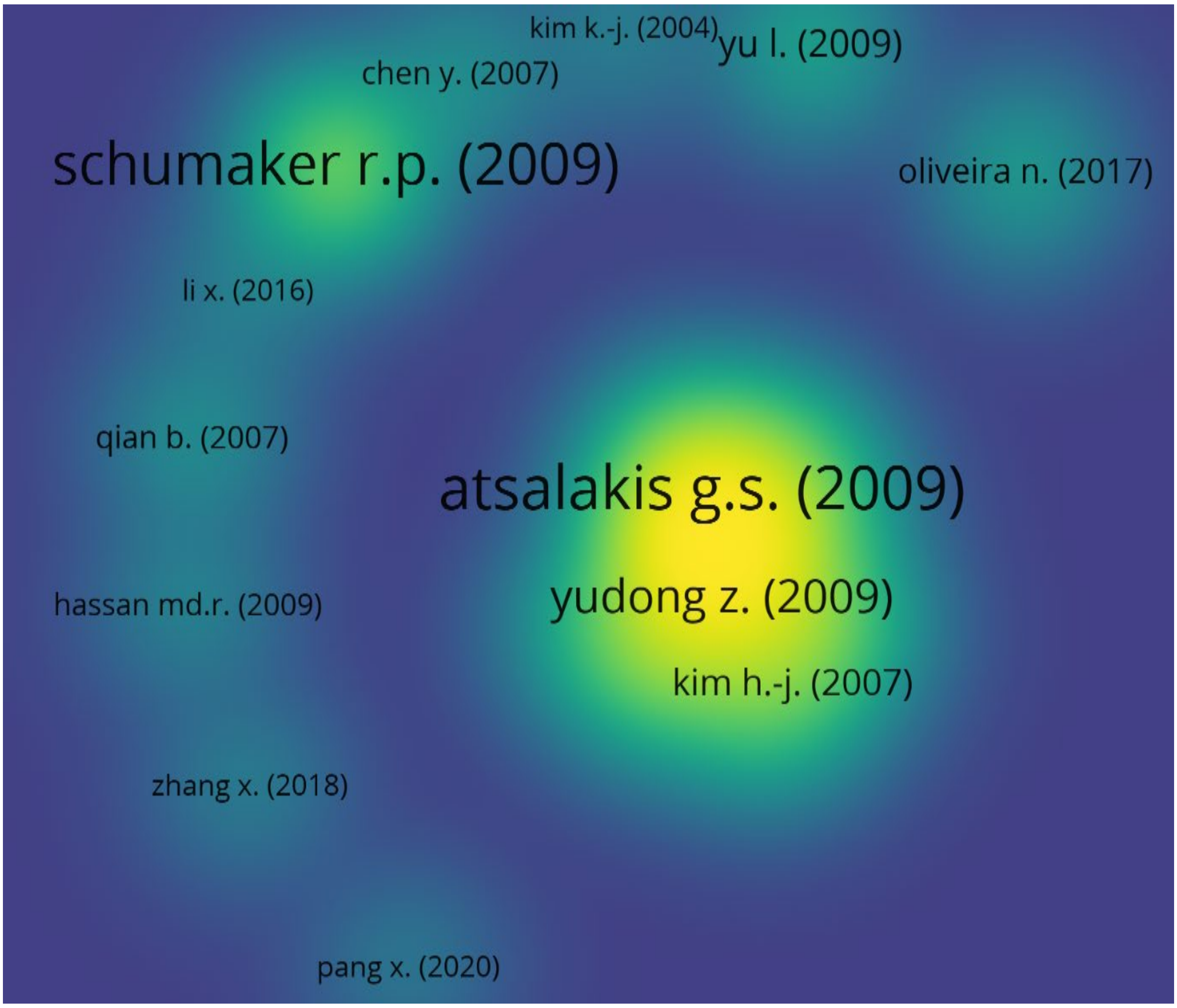

Fig. 6: Article citation network analysis.

Furthermore, Table 5 details the prominent research articles in SMF in terms of year of first publication (YFP), age, T.C., and links. The journals considered were from 2001 to 2021, over two decades. The top 5 leading articles were Atsalakis and Valavanis (2009), Schumaker and Chen (2009), Guresen et al. (2011), Yudong and Lenan (2009) and Ticknor (2013).

Table 5: Most cited research articles in SFM.

Author(s)	Focus of Study	YFP	AGE	TC	TC/AGE	links
Atsalakis and Valavanis (2009)	Survey of neural and neuro fuzzy techniques in SMF	2009	12	451	37.58	11
Schumaker and Chen (2009)	Forecasting of Stock market using textual analysis from breaking news	2009	12	421	35.08	0
Guresen et al. (2011)	Application of ANN in stock market index forecasting	2011	10	362	36.20	7
Yudong and Lenan (2009)	BCO and B.P. neural network to SMF	2009	12	271	22.58	4
Ticknor (2013)	SMF using Bayesian regularized ANN	2013	8	264	33.00	4
Boyacioglu and Avci (2010)	Better prediction of stock market using Adaptive Network-based Fuzzy Inference System	2010	11	250	22.73	3
Chong et al. (2017)	Analysing the stock market using deep learning techniques	2017	4	246	61.50	8
Hadavandi et al. (2010)	Prediction of stock market using ANN and genetic fuzzy systems	2010	11	243	22.09	5
Armano et al. (2005)	Prediction of stock market indexex hybrid genetic-neural model	2005	16	195	12.19	5
Hassan et al. (2007)	Integration of G.A., ANN and HNM in predicting the stock market	2007	14	193	13.79	2

Tsai and Hsiao (2010)	Multiple feature extraction for SMF	2010	11	182	16.55	2
Yu et al. (2009)	Application of least square SVM in stock market trend mining	2009	12	154	12.83	0
Kim and Shin (2007)	Integration of G.A. and neural networks in identifying stock market patterns	2007	14	143	10.21	4
Oliveira et al. (2017)	The effect of weblog data in SMF	2017	4	129	32.25	0
Göçken et al. (2016)	Application of Hybrid ANN models in SMF	2016	5	123	24.60	2
Majhi et al. (2009)	Development of FLANN model for predicting the stock market	2009	12	119	9.92	3
Asadi et al. (2012)	Hybrid intelligent modelling for stock market index forecasting	2012	9	97	10.78	3
Qian and Rasheed (2007)	Effect of multiple classifiers in stock market prediction	2007	14	94	6.71	0
Chen et al. (2007)	Stock market modelling using flexible neural trees	2007	14	92	6.57	0
Singh and Srivastava (2017)	Stock prediction using deep learning	2017	4	81	20.25	2

By carefully examining the analysis in Table 5, it is evident that Atsalakis and Valavanis (2009) was the most cited work in the league of top articles in the area of SMF. In Atsalakis and Valavanis (2009), the authors examined over 100 articles related to neuro-fuzzy techniques. They classified based on input data, forecasting method, performance assessment and performance measures. In Schumaker and Chen (2009), the authors investigated how breaking news and stock quotes directly impacted stock prices. The limitation to this work was that the modelling was done using a small data set. Therefore, there is a need to have a modified model that can accommodate large datasets and eliminate market biases associated with a short time window.

In Guresen et al. (2011), the authors studied the difference in prediction accuracy by comparing three sets of ANN models using the same NASDAQ stock exchange dataset. Their experiments showed that the Multi-layer perception ANN Model outperformed the other two models; dynamic ANN and hybrid N.N. (GARCH-ANN). In subsequent works like Armano et al. (2005), Asadi et al. (2012), Boyacioglu and Avci (2010), Göçken et al. (2016), Hadavandi et al. (2010), Hassan et al. (2007), Kim and Shin (2007), Ticknor (2013), Tsai and Hsiao (2010), and Yudong and Lenan (2009); hybridisation, integration, combination, ensembling, fusion, and adaptive strategy of two or more techniques were employed to develop more robust models that can handle the volatility and chaotic nature of stock data. Consequently, by measuring the prediction accuracy, integrated strategies outperform individual models in most of the research conducted.

3.6. Influential countries in SMF research

Country-wise, some nations actively contribute to promoting research on SMF than others. Even though these can be attributed to many factors, different countries must engage in joint research to promote research outputs. The success of these countries can be attributed to scholarships, fellowships, funding, grants and much more that the constituted authorities of such countries made available to support research excellence in their respective nations. Fig. 7 below illustrates the countries with the most contributions and productivity in SMF. It is worth noting that this analysis is not motivated for any political gain or research supremacy but purely based on research outputs and impacts that the countries have made towards remarkable landmarks in SMF research. Fig. 7 was generated by setting the software to plot a map using a threshold of 5 documents and 50 citations. Out of 47 countries, only 15 met the defined threshold.

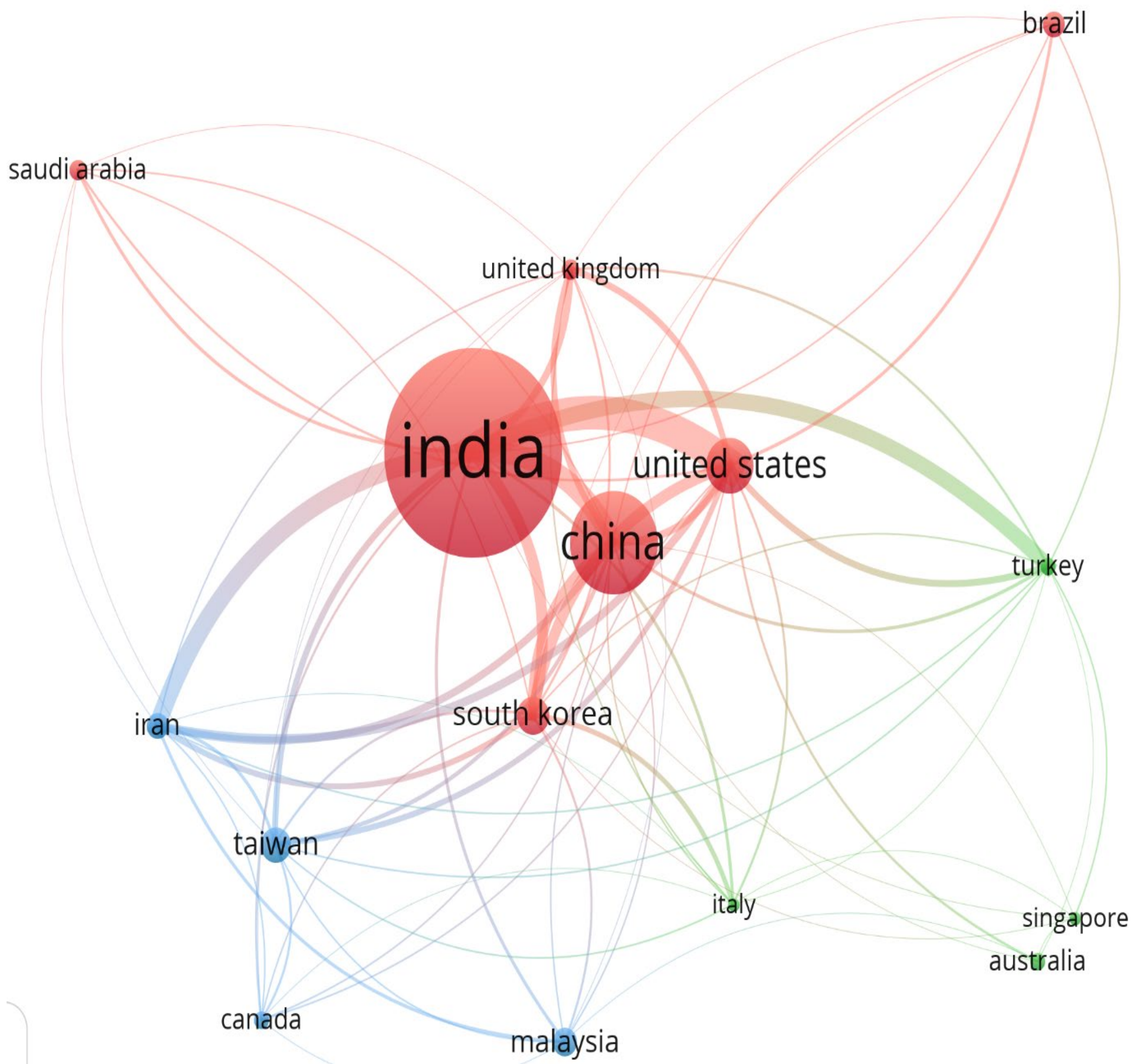

Fig. 7: Network of influential countries in SMF research.

Out of 193 United Nations recognised countries, only 47 nations are involved in SMF, which is equivalent to 24%. Out of the 47, only 15 met the stated threshold. The respective sizes of the nodes depict each nation's contribution to the development of prediction techniques to forecast the stock markets. For instance, with 71 articles, India appears to be the biggest and the most productive nation in this area, followed by China (35), the USA (19), South Korea (13) and Taiwan (12). In contrast, referring to the analysis in Table 6, in regard to angle of citations, which is a measure of impacts, the USA appears to be the most impactful country in this category with the highest citations, i.e. 1792. After the USA, the other rankings in a descending order are China (1070), South Korea (758), Turkey (737), and India (626).

Table 6: Most influential countries in SMF research.

Country	Continent	No. of Articles	TC	AC	TLS
India	Asia	71	626	9	152
China	Asia	35	1070	31	82
United States	North America	19	1792	94	111
South Korea	Asia	13	758	58	69
Taiwan	Asia	12	348	29	38
Malaysia	Asia	10	62	6	20
Iran	Asia	9	492	55	65
Brazil	South America	9	88	10	12
United Kingdom	Europe	7	434	62	39
Saudi Arabia	Asia	7	67	10	14
Turkey	Europe	6	737	123	46
Australia	Oceania	6	335	56	7
Canada	North America	6	75	13	17
Italy	Europe	5	272	54	23.0
Singapore	Asia	5	70	14	7

India is the most productive country in this area and the most cross cited country. Globally, not all the continents are represented in the mapping analysis. As presented in Table 6, the least contributing continent is Oceania, while Asia has the most contribution. Asia is actively participating in this research, as 8 out of the 15 listed countries are effectively and efficiently pooling efforts and resources to explore SMF research. However, it is worth noting that no country in Africa is actively involved in research related to SMF.

3.7. Subject area distribution

It is essential that we analyse the discipline roles and subject areas that have contributed to the stock market prediction techniques. Based on search algorithm results on Scopus, 10 distinct disciplines have contributed most to this research area. The topmost contributing discipline is computer science, with a total of 165 articles, which accounts for a total percentage of 38.4% of all the subject areas combined. Engineering is second on the list with 83 articles, mathematics third with 45, while Business Management and Accounting and Decision Science published 29 and 27 documents, respectively. Fig. 8 illustrates the percentage contributions of the leading disciplines in the area of SMF. At the same time, Table 7 shows the frequency of publication of the top 5 leading disciplines in SMF.

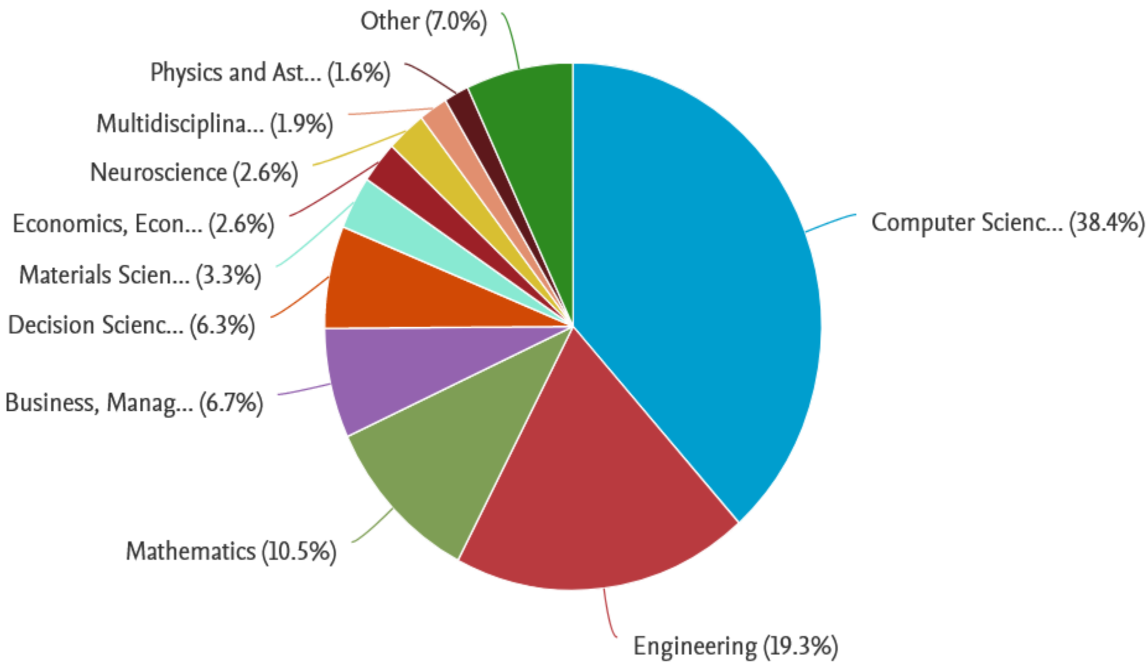

Fig. 8: Subject area distribution in SMF research.

Table 7: Disciplines participation in SMF research.

Subject Area	**No. of articles**
Computer Science	165
Engineering	83
Mathematics	45
Business, Management and Accounting	29
Decision Sciences	27

3.8. Major stock market research areas

It is vital to analyse keywords in order to understand key research areas (Gupta & Dhawan, 2018). Keywords describe the domain within which a particular research area is concentrated. The title and abstract primarily reflect the content of the research article. Holistically, analysing the keywords reveal the patterns and trends in Stock Market Research Area. Keywords in any article are mostly aligned with the research theme and content. Some researchers claim that keywords are essential to identify patterns and trends in a given subject, so mapping keywords provides a complete understanding of the subject of interest (Chen & Song, 2017; Su et al., 2010). Although the keyword co-occurrence network was discussed earlier, only the author keywords were used. This section offers keywords that cut across the title and abstracts of the entire 220 downloaded journals. For retrieval, the default minimum number of occurrences was 10. However, following the recommendations of conducting multiple experiments to get optimal graphic (Yin et al., 2019), 15 appears to be an appropriate metric to adopt.

The counting was set as binary based on standard practice, while the number of occurrences was put at 15. Out of 4838 terms, only 73 met this threshold. Further, the default benchmark of 60% of the most relevant terms was applied to ensure that the selected words were highly relevant. Therefore, this leads to a drop of captured items to 44. The key areas that need attention based on co-occurrence analysis are shown in Fig. 9. According to Fig. 9, there are three clusters: red, green and blue. The visualisation map shows "market" as the central theme. Essential keywords in the red cluster emphasise machine learning techniques used to analyse trends and patterns in SMF research. Also, significant keywords in the green cluster contain terminologies that apply to deep learning tools in predicting the stock market. Lastly, the blue cluster indicates the input, parameters, indices and common markets used to conduct stock market research. The essential keywords in all the clusters measure the application of Artificial Intelligence (A.I.) tools in predicting the stock market. Meanwhile, machine learning seems outdated to most researchers, as many of the latest research uses sophisticated deep learning tools, which are more robust than the existing machine learning models.

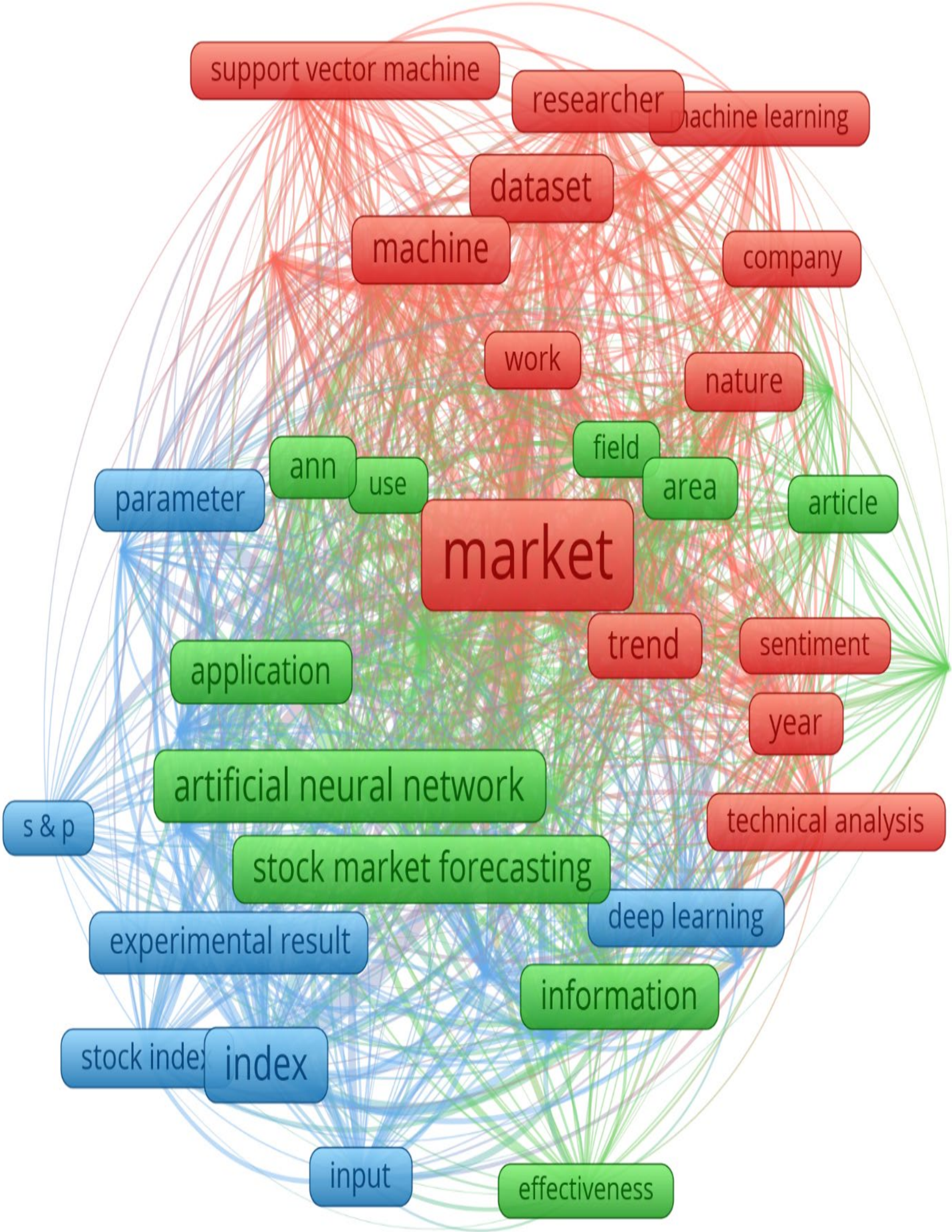

Fig. 9: Clusters of SMF research keywords.

Table 8 shows the most recurring keywords in the title and abstracts of the articles used in conducting reviews. The analysis shows the frequency of occurrences (O), and at the same time, computes their relevance score (R.S.) to the subject area.

Table 8: Most recurring words in SMF research articles.

Term	O	RS
Market	132	0.5354
Index	59	1.3561
Artificial Neural Network	55	1.2898
Problem	49	0.5677
Stock Market Forecasting	47	0.4878
Machine	45	0.6444
Dataset	43	0.6536
Trend	41	1.1461
Information	41	0.6533
Day	38	0.7488
Application	38	0.6494
Technical Indicator	36	0.6006
Experiment	35	0.6291
Parameter	34	0.51
Ann	32	1.4177
Area	32	0.7936
Year	32	0.7602
Time	32	0.5313
Time Series	32	0.4602
Experimental Result	31	1.0905

4. Discussion of the findings.

Prediction of stock market trends is a research area that cut across several scientific fields such as computer science, finance, engineering, mathematics and many more. Efficient Market Hypothesis (EMH) presumes that a share price is a reflection of the entire market information (Ţiţan, 2015). However, researchers have debunked this controversial idea, and subsequently embraced the theories of fundamental analysis (fair value indices) and technical analysis (trends and charts). Subsequently, the technical analysis has been dramatically transformed into the features explored in building machine learning and deep learning models. With AI, market forecasting often comes as either a classification problem or regression problem, using tools like logistic regression and support vector machine (Alpaydin, 2014). The key focus areas in SMF is analysed under the following techniques.

4.1. Prediction techniques

Artificial Neural Network, Decision Support System, Hidden Markov Model, Recurrent Neural Network, Naïve Bayes, Support Vector Machine, and many more are popular techniques used to accurately predict stock performance rather than the traditional statistical methods. These approaches have robust methods of capturing the performance and the determinant factors of any financial market. In Ticknor (2013), the stock price was predicted using the Bayesian Regularized Artificial Neural Network. The investigators made use of technical indicators and daily stock prices as inputs in predicting a future price. The outcome of their experiment showed that building an efficient and effective prediction model is possible without undergoing the complex tasks of data preprocessing and seasonality testing.

In Shrivas and Sharma (2018), the investigators integrate support vector machine, Chi-square Automatic Interaction Detector, Classification and Regression Technique (CART), and ANN to predict the financial market. The results of their experiment give Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) of 0.0044 and 0.676, respectively. Also, in Chen and Hao (2017), Feature Weighted Support Vector Machine and Feature Weighted K-Nearest Neighbour were hybridised to build a framework to predict two Chinese markets. The result of their proposed model confirmed that feature weighted techniques are more robust than their counterparts. Similarly, in Zhong and Enke (2017), the stock market daily direction was evaluated by combining ANN with any of the Principal Component Analysis (PCA), Kernel-Based Principal Component Analysis (KPCA), and Fuzzy Robust Principal Component Analysis (FRPCA). Based on their simulation results, it was deduced that the combination of ANN with PCA gives higher risk-adjusted profit than other combinations. Further, they pointed out that delicate selection of kernel function can give better results in terms of ANN combined with KPCA. Relatedly, Hadavandi et al. (2010) applied the integration approach by combining ANN and Genetic Fuzzy System (GFS) to design a stock price prediction model. The hybrid intelligence model giave a better forecast, judging by its MAPE, when compared with the existing forecasting models.

Recently, renowned deep learning techniques have taken over in predicting the stock market trends. Some studies have queried the replicability and reproducibility of machine learning methods (Makridakis et al., 2018); hence, these limitations have led to a rise in the usage of deep learning techniques. Many researchers have deeply explored deep learning models and their hybridisation in making more accurate stock market predictions. For instance, in Kara et al. (2011), the Istanbul stock exchange market was predicted using ANN and SVM models; their findings showed that prediction accuracy was 75.74% and 71.52%, respectively. Therefore, the result is one of the proofs that deep learning models are always robust and efficient in predicting the market compared to existing statistical or machine learning models. Deep learning applications cut across several modelling such as object classification (Zhao et al., 2019), image classification (Jiang & Zhang, 2020; Rawat & Wang, 2017) and time series forecasting (Brownlee, 2018; Jiang & Zhang, 2018). Deep learning can handle big data, discover the non-linear association between the input variables and prediction outputs. In many cases, many experiments have shown that deep learning models outperform linear and machine learning models.

4.2. Clustering techniques

Many researchers have used the filtering approach in forecasting the stock market. Over the years, researchers have investigated different applications of computational intelligence in solving financial market problems. In Cavalcante et al. (2016), the authors reviewed studies covering preprocessing and clustering of market data. In Arévalo et al. (2017), flag pattern trading rules were developed to stop losses and take profits in short and medium trade. The developed model was able to limit extreme loss in each trade to 100 points. Similarly, in Srinivasan and Ibrahim (2010), using intraday data, the authors predict the conditional variance of SENSEX index market returns using a GARCH model.

Apart from filtering, another clustering technique commonly used is Fuzzy Logic (F.L.), which is a computing technique based on the level of truth. Instead of the regular Boolean Logic of using “1” or “0” to denotes “true” or “false”, Fuzzy logic purports that the truth value of a variable is any real number between 0 and 1. In Wei et al. (2011), the authors created an advanced forecasting model based on ANFIS, utilising several technical indicators to predict the market trend. However, the limitation in their work was that the model lacked robustness in predicting other stock indices apart from the Taiwan Stock Exchange (TAIEX).

In building an investment portfolio, the essence of choosing an effective classifying technique cannot be overemphasised. An efficient portfolio management system is achievable using appropriate clustering techniques, which helps in the grouping of similar stocks to identify the best performing one. In Nanda et al. (2010), a data mining approach was used in classifying stocks. The validation of their experiment showed that k-clusters outperformed the other two clusters; Fuzzy C-means and self-organising maps.

4.3 DEA and Machine Learning Techniquess

Though machine learning algorithm has been extensively used in recent years, it is still challenging that it fails to evaluate the efficiency of stocks before predicting their returns. In addition, machine learning has difficulty telling between performing and non-performing stocks. Data Envelopment Analysis (DEA) is a popular tool for dealing with this issue, especially when many inputs and outputs characteristics of multiple equities are being considered. In recent times, many researchers now leverage the integration of DEA with Machine Learning algorithms to build super robust models that are computationally fast and accurate. Zhu (2020) suggested using the DEA technique to speed up the processing time in a huge data scenario. The authors suggested DEA as a data-driven analysis tool for evaluation and benchmark testing. Even though traditional DEA has established methods to calculate large-scale data (decisions made, units, inputs, and outputs), the precious knowledge of the network structure hiding in big data can be recovered by DEA (Zhu, 2020). For stochastic dynamic facility layout challenges, Tayal et al. (2020) used meta-heuristics, DEA, and machine learning methods to evaluate efficiency. They suggested a three-step process to acquire the distinct ranks and estimate the efficiency values of the layouts, with DEA boosted by supervised and unsupervised ML approaches.

The redevelopment of city blocks was studied by Qu et al. (2019), who also presented a DEA, deep learning model. They presented the DEA and used a deep learning model to foretell output-oriented metrics of non-intensive blocks. Integration of DEA with clustering techniques was used to increase precision by Mirmozaffari et al. (2020). Aydin and Yurdakul (2020) used a DEA method with four ML algorithms (k-means, hierarchical clustering, decision tree, and random forest) to assess how well 142 countries fared against COVID-19 in a new three-staged framework. Using machine learning approaches, Rebai et al. (2020) were able to make accurate predictions about the performance of secondary schools in Tunisia. A two-stage analysis procedure was implemented. The directional distance function (DDF), estimated by the data envelopment analysis (DEA), was used to handle the desired outputs in the first stage, while the decision tree and random forest algorithms were employed in the second stage to identify and depict the high school performance variables. A petrochemical plant adaptive capacity was examined and enhanced by Salehi et al. (2020). Indicators of resilience engineering, including redundancy and cooperation, were computed and analyzed using the DEA technique. They used a multilayer perceptron (MLP) to analyze the available data and generate an assessment of the system's adaptive capability.

To foresee the effect of environmental factors on farm performance, Nandy and Singh (2020) analyzed farm productivity using machine learning and DEA methods. With the use of ML approaches, Jomthanachai et al. (2021) identified a set of risk variables via FMEA and predicted the level of residual risk based on simulated data matching the risk treatment scenario. The effectiveness of corporate governance and financial institutions was evaluated by Thaker et al. (2022), who combined random forest regression with DEA in a two-stage model. Using BPNN-DEA, GANN-DEA, ISVM-DEA, and SVM-DEA, Zhu et al. (2021) coupled machine learning with DEA. The effectiveness of China's manufacturing firms was measured, and projections were made. By combining DEA models and ML methods, Jomthanachai et al. (2021) analyzed risk management. To make their risk assessment, they combined DEA with ANN for cross-efficiency. In order to combine DEA with machine learning, Zhong et al. (2021) implemented a Super SBM-DEA-BPNN method. The authors' ultimate goal is to boost fusion efficiency. Taherinezhad and Alinezhad (2022) employed DEAML integration to estimate country efficiency scores from the COVID-19 data. The regression goal was predicted using GANN, BPNN, SVM, and ISVM (efficiency values). Farm productivity was calculated by Nandy and Singh (2020) using a combination of machine learning and data engineering. They employed a two-step DEA model built on random forest and logistic regression. In sum, an integrated ML-DEA is an open area to explore by upcoming researchers in this area of study.

4.4 Inverse optimisation

Inverse optimisation is a mathematical technique for determining optimal changes in the data parameters of a system, such that the initial optimal solution of the system is always preserved or unchanged. To explain its underlying principle, consider a system with x and y to be its input and output parameters, respectively. Let θ_1 denote the optimal solution of the system at time $t = 1$. Suppose, at time $t = 2$, the output parameter, y, changes to $y + \Delta y$, to give an optimal solution, θ_2. If Δy is known, we want to determine the corresponding change in input parameter, Δx, that will give a new input, $x + \Delta x$, such that $\theta_1 = \theta_2$ (i.e. the optimal solution of the system is retained). This summarises the evaluation mechanism, Inverse DEA model belongs to the family of inverse optimisation techniques. Inverse optimisation can be solved using column generation and ellipsoid methods. The column generation method solves the initial problem by generating the required columns for applying the simplex method. The ellipsoid method solves inverse problems with a polynomial-order algorithm. Inverse optimisation can be applied to the stock market to forecast the returns of multiple stocks simultaneously.

4.5 Inverse DEA (I-DEA)

Another less computational area that needs to be explored in stock market forecasting is Inverse DEA. Inverse problems have only been created and researched in recent years. There are two main types of models in this field, and they are the ones that address issues of resource allocation and those that address issues of investment analysis. To find the optimal input for a given output while maintaining the existing efficiency value of the DMU under assessment relative to other DMUs, the DEA resource allocation problem becomes an inverse DEA (I-DEA) problem. Investment analysis problems are another sort of IDEA model. Its goal is to find the most beneficial outcome feasible for a given input while maintaining the same efficiency value of DMUo relative to other DMUs. An I-DEA model for input and output estimates was first proposed by Wei et al. (2000). How much more output could a DMU produce from a given input if it maintained its current efficiency value relative to other DMUs in the group? Alternately, how much more input may be given to the unit if the outputs are increased to a specific number and the efficiency of the unit remains the same? The I-DEA model was turned into multi-objective linear programming (MOLP) problem, with the increment in inputs and outputs assumed to be positive. Yan et al. (2002) introduced a preference-cone constrained I-DEA problem that reflected decision-maker preferences and was helpful in resource planning. Hadi-Vencheh and Foroughi (2006) proposed an expanded version of I-DEA model that incorporate the effects of increasing some inputs (outputs) while decreasing inputs (outputs) of other DMUs simultaneously. Using DEA and MOLP, the authors indicated that the approach presented by Wei et al. (2000) did not ensure the application of the analysis since each DMU might lead to an increase in some inputs (outputs) and a decrease in other inputs (outputs). The projection evaluation of stock return based on Inverse DEA is rarely discussed in the literature, and this make it a potential hotspot for researchers to explore.

5. Knowledge gaps and potential research directions

Science mapping is a network visualisation technique that mainly analyses literature using bibliometric or scientometric approaches (Hosseini et al., 2018). This research considers a scientometric approach to reviewing pieces of literature on SMF. It extends the previous integral and critical studies by applying a quantitative mechanism of mapping out past works on SMF using VOSviewer software. Apparently, there are shortcomings in this body of literature. Machine learning tools are noticeably outdated and many works of literature also explored different deep learning tools, especially ANN. In recent years, ANN has been used extensively in business (Tkáč & Verner, 2016). Significantly, the application of ANN to forecast the market index has yielded promising results with less MAE and MAPE. However, more is needed to explore fusion and hybridisation to have a significant breakthrough in predicting an ever dynamic and volatile market, especially on a global level.

Despite being evident that significant progress has been made in stock market prediction, there still exists room for improvement. Most of the previous studies consider one market in their analysis. However, due to the differences in stock market policies, it has been challenging to test existing models on other markets. With this, there is a need to build more models that can perform cross-market analysis prediction. More recently, in Hoseinzade and Haratizadeh (2019), Hoseinzade et al. (2019), Lee et al. (2019), Nguyen and Yoon (2019), attempts were made for prediction using several combined markets. Noticeably, the best among them was Lee et al. (2019), as the authors validate their model using several markets in 31 different countries, thus yielding good results. It is worth noting that there is a limitation to this model, as the leading market used is a big market in a developed country, and the model replication and validation is mainly conducted on small markets in emerging countries. Efforts should be raised to develop better models that can accurately forecast stock market returns, especially in an unprecedented economic collapse such as in the Covid-19 pandemic. Although there have been significant breakthroughs in addressing stock market predictions using different neural networks, more strategic and collaborative methods should be adopted to use a less computational and less complex model that outperform the existing models.

Another major challenge identified is the volatility in the market system, which affects the overall performance of the global market. Previous works have examined or forecasted stock market indices, but most of their analysis was focused on a specific market or stock. The use of earlier techniques might be inadequate and cumbersome for those investors who wish to make intelligent decisions about which stock market or stock to invest in, because an analyst must assess the prospects of each market or stock individually. In other words, the problem takes a new dimension when considering that an analyst would have determined between performing and non-performing stocks. To better assess the efficiency of related stocks and make reliable projections for returns, more efforts should be pooled to adopt a modified technique of existing models like Data Envelopment Analysis (DEA) and Inverse DEA.

The DEA model explicitly advocates evaluating the relative performance or efficiency of homogenous groups of decision-making units (DMU) that provide a similar type of service. DEA promises a more accurate and robust method of comparing market indices on a global scale. A market's efficiency score can be used as a decision-making tool by an investor to select the best market or DMU. In addition, past and recent modifications of the DEA model have proven beyond doubt that it can perform more than a performance evaluation function, as it can be configured or extended to accurately forecast the stock market indices. However, the DEA methodology requires an analyst to carefully select the correct combination of inputs and outputs so that results are accurate. As a first rule, when applying the DEA model to performance evaluation of the global stock market, the selected individual markets must possess the same type and number of input and output parameters to avoid any bias. Homogeneity in DEA applications is based on this first rule.

Furthermore, using ratios or percentages to express input and output quantities is a pitfall that should be avoided when using the DEA model. As there is an identical set of inputs and outputs for each market, no conversion is required since the DEA model can handle different dimensions' inputs and outputs without dimensional conflicts. By solving the linear program created by DEA, each stock market is given an efficiency score. In other words, the efficiency score measures the performance of each market.

If the efficiency score is one, the evaluated market is efficient, while if it is less than one, the market is inefficient. In this way, the efficient markets will now effectively serve as the benchmarks for the inefficient ones. In this connection, an investor can easily choose from the set of efficient markets or the benchmarks identified by the model. An investor can decide when to invest in an inefficient market, based on the possibility for gains in the future. Such markets require accurate forecasts based on a modified DEA model. In order to accomplish this, the DEA model can be modified to use the efficiency scores of each inefficient market along with the inputs and outputs from each efficient market to generate its output values. Thus, it becomes possible to forecast market retuurns for selected inefficient markets. This technique is promising because DEA has an excellent reputation as a benchmarking and management tool in the field of operations research and econometrics.

A second option would be to consider evaluating the performance of one stock market over a specific planning period in years. This case considers the specified years as decision-making time units (DMTU) in the DEA analysis. Similarly to our DMU evaluation, the standard DEA will calculate an efficiency score per year, followed by an assessment of whether a year was efficient or inefficient. Based on this knowledge, analysts can identify the year(s) in which the stock market did well or poorly and use it to forecast market indices in later years. It is important to note that while this is a brief analysis of another DEA methodology, real-life scenarios require an analyst to carefully select the best trading years of the chosen market as a reliable basis for future forecasting. In the forecasting process, accuracy will be significantly affected by choice of the chosen years or decision-making time units (DMTU).

As a final consideration, the DEA model may be used to determine potential improvements for an inefficient stock market compared with efficient ones on a global scale. The literature has yet to address this exciting research gap. As DEA models are used to evaluate the performance of individual markets, it is crucial to determine the optimal level of inputs and outputs for inefficient markets that will place them on par with efficient markets. In the DEA methodology, this can be accomplished by using a dual form of the linear program. Using this information, stockbrokers and market experts can control cash inflows and outflows into and out of the market. This would, however, require some understanding of microeconomics and linear programming.

When the DEA is applied to complex problems in the real world, the only challenge is the modification or extension necessary to accommodate a particular problem. A thorough understanding of the real world coupled with good mathematical skills in linear programming is necessary in most cases. In addition, the modification or extension of the DEA model must be feasible and produce global optimal results for the forecasts to be accurate. In view of the numerous high-impact publications and research works in the field of stock market prediction and analytics, it is recommended that future work adopt a more robust approach that involves the use of extended DEA models in dual or extended forms, to evaluate performance and forecast stock market returns.

6. Conclusion

There is a growing concern about building sophisticated models that can analyse and forecast the market better. Over the years, ANN has been extensively used. However, due to the dynamic and volatile nature of the stock market, there is a need to develop more robust and sophisticated models that can accurately forecast the global market, especially in the presence of uncertainty. This study presents the first scientometric review and survey of stock market prediction techniques, in which 220 highly reputable journals were assessed using science mapping software. This study conducts a survey and scientometric analysis of the stock market prediction techniques over two decades (2001-2021).

A single robust database, i.e., Scopus, was used during the analysis to retrieve bibliometric data of 220 top journals. The assessment reveals that the earliest significant rise in this study area was recorded in 2007. Even though the first document from the databank was in 2001, researchers have channeled more efforts into this research area, which caused a surge in the research output, especially in recent years. In 2019, the number of documents published per year reached its all-time high of 42 documents per annum. After, there was a steep decline in 2020, owing to the unprecedented event of the Covid-19 pandemic,

Our study observed that the most collaborative strength involved three researchers, namely Adekoya, Weyori, and Nti. Also, another analysis showed that “expert systems with applications” is the most influential and the most cited journal outlet, with 2676 citations. Based on the frequency of author keywords, we found out that stock market prediction, stock market, machine learning, deep learning, and stock market forecasting were the most used keywords by previous investigators. From the 220 articles reviewed, it can be deduced that ANN’s and Fuzzy-Based techniques are the most adopted prediction and clustering techniques, respectively.

Analysis of geographical areas shows that out of 193 United Nations recognised countries, only 47 of them conducted research relating to the prediction of the stock market, which is equivalent to 24%. In terms of productivity, the leading nations were India (71), China (35), the USA (19), South Korea (13) and Taiwan (12). In contrast, in terms of Impacts, the top 5 leading countries were the USA (1792), China (1070), Korea (758), Turkey (737), and India (626).

Also, while mapping out relevant keywords from paper titles and abstracts, we found that technology has shifted from using machine learning models to sophisticated models like deep learning and hybridised models. The question faced by SMF is that historical data and data mining are usually affected by certain factors such as government policy, market sentiments and many others. Thus, there is a need to build up a more robust model that can handle this problem and still predict better than existing models. In this realization, this study suggests a less computational approach, Inverse DEA, as an alternative to forecasting stock returns. More so, this study suggests that researchers could also explore hybrid methodologies of DEA and machine learning algorithms in building more sophisticated models that can handle the challenges of the stock market system.

It is worth noting that the outcomes of this review were analysed under certain constraints. Foremost, the authors could not finish the manual reading of the entire context of all the papers retrieved from the data bank; thus, there might have been one or two crucial pointers that are not emphasised in this work. Nevertheless, the authors tried their possible best to ensure adequate coverage of all related articles. Second, the scope of this review is limited to SMF, without discussing the entire financial markets. In conclusion, this study highlights and summarises the efforts, research and progress made in SMF. This article is helpful to many stock market participants and organizations, such as research students, industry, government, policymakers, analysts, investors, managers, and profitable organisations.

Declaration of competing interest

The authors declare that they have no known financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The work described in this paper was supported by The Hong Kong Polytechnic University under student account code RLN7. The authors also would like to thank The Hong Kong Polytechnic University Research Committee for financial and technical support.

References

Aghaei Chadegani, A., Salehi, H., Yunus, M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian social science*, *9*(5), 18-26.

Alpaydin, E. (2014). Introduction to machine learning (adaptive computation and machine learning series). In: MIT Press Cambridge.

Araújo, R. d. A., & Ferreira, T. A. (2013). A morphological-rank-linear evolutionary method for stock market prediction. *Information Sciences*, *237*, 3-17.

Arévalo, R., García, J., Guijarro, F., & Peris, A. (2017). A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. *Expert Systems with Applications*, *81*, 177-192.

Armano, G., Marchesi, M., & Murru, A. (2005). A hybrid genetic-neural architecture for stock indexes forecasting [Article]. *Information Sciences*, *170*(1), 3-33. https://doi.org/10.1016/j.ins.2003.03.023

Asadi, S., Hadavandi, E., Mehmanpazir, F., & Nakhostin, M. M. (2012). Hybridization of evolutionary Levenberg-Marquardt neural networks and data pre-processing for stock market prediction [Article]. *Knowledge-Based Systems*, *35*, 245-258. https://doi.org/10.1016/j.knosys.2012.05.003

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques - Part II: Soft computing methods [Article]. *Expert Systems with Applications*, *36*(3 PART 2), 5932-5941. https://doi.org/10.1016/j.eswa.2008.07.006

Aydin, N., & Yurdakul, G. (2020). Assessing countries' performances against COVID-19 via WSIDEA and machine learning algorithms. *Applied Soft Computing*, *97*, 106792.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. Third international AAAI conference on weblogs and social media,

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*.

Blažun, H., Kokol, P., & Vošner, J. (2015). Research literature production on nursing competences from 1981 till 2012: A bibliometric snapshot. *Nurse Education Today*, *35*(5), 673-679.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327.

Boyacioglu, M. A., & Avci, D. (2010). An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange [Article]. *Expert Systems with Applications*, *37*(12), 7908-7912. https://doi.org/10.1016/j.eswa.2010.04.045

Brownlee, J. (2018). *Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A Systematic review. *Expert Systems with Applications*, *156*, 113464.

Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics*, *66*(2), 327-343.

Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., & Oliveira, A. L. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, *55*, 194-211.

Chen, C. (2017). Science Mapping: A Systematic Review of the Literature. *Journal of Data and Information Science*, *2*(2), 1-40. https://doi.org/doi:10.1515/jdis-2017-0006

Chen, C., & Song, M. (2017). Measuring Scholarly Impact. In *Representing Scientific Knowledge* (pp. 139-204). Springer.

Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, *80*, 340-355.

Chen, Y., Yang, B., & Abraham, A. (2007). Flexible neural trees ensemble for stock index modeling [Article]. *Neurocomputing*, *70*(4-6), 697-703. https://doi.org/10.1016/j.neucom.2006.10.005

Cheng, C.-H., Chen, T.-L., & Wei, L.-Y. (2010). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Information Sciences*, *180*(9), 1610-1629.

Cheng, P., Quek, C., & Mah, M. L. (2007). Predicting the impact of anticipatory action on U.S. Stock Market - An event study using ANFIS (A Neural Fuzzy Model) [Article]. *Computational Intelligence*, *23*(2), 117-141. https://doi.org/10.1111/j.1467-8640.2007.00304.x

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

Cobo, M. J., López‐Herrera, A. G., Herrera‐Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for information Science and Technology*, *62*(7), 1382-1402.

Cobo, M. J., López‐Herrera, A. G., Herrera‐Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for information Science and Technology*, *63*(8), 1609-1630.

Fabregat-Aibar, L., Barberà-Mariné, M. G., Terceño, A., & Pié, L. (2019). A bibliometric and visualization analysis of socially responsible funds. *Sustainability (Switzerland)*, *11*(9), 2526.

Ferreira, F. G., Gandomi, A. H., & Cardoso, R. T. (2021). Artificial intelligence applied to stock market trading: A review. *IEEE Access*, *9*, 30898-30917.

Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques [Review]. *Computer Science Review*, *34*, Article 100190. https://doi.org/10.1016/j.cosrev.2019.08.001

Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, *30*(2), 119-145.

Gaviria-Marin, M., Merigo, J. M., & Popa, S. (2018). Twenty years of the Journal of Knowledge Management: A bibliometric analysis. *Journal of Knowledge Management*.

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, *40*(16), 6266-6282.

Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research* (pp. 257-276). Springer.

Göçken, M., Özçalici, M., Boru, A., & Dosdoɨru, A. T. (2016). Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction [Article]. *Expert Systems with Applications*, *44*, 320-331. https://doi.org/10.1016/j.eswa.2015.09.029

Gupta, B. M., & Dhawan, S. M. (2018). Artificial Intelligence Research in India: A Scientometric Assessment of Publications Output during 2007-16. *DESIDOC Journal of Library & Information Technology*, *38*(6).

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

Guz, A. N., & Rushchitsky, J. (2009). Scopus: A system for the evaluation of scientific journals. *International Applied Mechanics*, *45*(4), 351.

Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting [Article]. *Knowledge-Based Systems*, *23*(8), 800-808. https://doi.org/10.1016/j.knosys.2010.05.004

Hadi-Vencheh, A., & Foroughi, A. A. (2006). A generalized DEA model for inputs/outputs estimation. *Mathematical and Computer Modelling*, *43*(5-6), 447-457.

Hassan, M. K., Aliyu, S., Saiti, B., & Halim, Z. A. (2020). A review of Islamic stock market, growth and real-estate finance literature. *International Journal of Emerging Markets*.

Hassan, M. R. (2009). A combination of hidden Markov model and fuzzy model for stock market forecasting. *Neurocomputing*, *72*(16-18), 3439-3446.

Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting [Article]. *Expert Systems with Applications*, *33*(1), 171-180. https://doi.org/10.1016/j.eswa.2006.04.007

Hjørland, B. (2013). Citation analysis: A social and dynamic approach to knowledge organization. *Information Processing & Management*, *49*(6), 1313-1325.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291-314.

Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables [Article]. *Expert Systems with Applications*, *129*, 273-285. https://doi.org/10.1016/j.eswa.2019.03.029

Hoseinzade, E., Haratizadeh, S., & Khoeini, A. (2019). U-cnnpred: A universal cnn-based predictor for stock markets. *arXiv preprint arXiv:1911.12540*.

Hosseini, M. R., Martek, I., Zavadskas, E. K., Aibinu, A. A., Arashpour, M., & Chileshe, N. (2018). Critical evaluation of off-site construction research: A Scientometric analysis. *Automation in Construction*, *87*, 235-247.

Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, *32*(10), 2513-2522.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Ibidapo, I., Adebiyi, A., & Okesola, O. (2017). Soft computing techniques for stock market prediction: A literature survey. *Covenant Journal of Informatics & Communication Technology*, *5*(2), 1-28.

Jabbar Alkubaisi, G. A. A., Kamaruddin, S. S., & Husni, H. (2017). A systematic review on the relationship between stock market prediction model using sentiment analysis on twitter based on machine learning method and features selection [Review]. *Journal of Theoretical and Applied Information Technology*, *95*(24), 6924-6933.

Jiang, W. (2021). Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications*, 115537.

Jiang, W., & Zhang, L. (2018). Geospatial data to images: A deep-learning framework for traffic forecasting. *Tsinghua Science and Technology*, *24*(1), 52-64.

Jiang, W., & Zhang, L. (2020). Edge-siamnet and edge-triplenet: New deep learning models for handwritten numeral recognition. *IEICE Transactions on Information and Systems*, *103*(3), 720-723.

Jin, R., Gao, S., Cheshmehzangi, A., & Aboagye-Nimo, E. (2018). A holistic review of off-site construction literature published between 2008 and 2018. *Journal of Cleaner Production*, *202*, 1202-1219.

Jomthanachai, S., Wong, W.-P., & Lim, C.-P. (2021). An application of data envelopment analysis and machine learning approach to risk management. *IEEE Access*, *9*, 85978-85994.

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, *38*(5), 5311-5319.

Katie, K., Justyna, P., & Ali, M. (2021). *2021 Capital Markets Fact Book*. https://www.sifma.org/resources/research/fact-book/

Kim, H. J., & Shin, K. S. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets [Article]. *Applied Soft Computing Journal*, *7*(2), 569-576. https://doi.org/10.1016/j.asoc.2006.03.004

Kim, K.-j., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, *19*(2), 125-132.

Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990). Stock market prediction system with modular neural networks. 1990 IJCNN international joint conference on neural networks,

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2021). A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting. *Annals of Data Science*, 1-26.

Lee, J., Kim, R., Koh, Y., & Kang, J. (2019). Global stock market prediction based on stock chart images using deep q-network [Article]. *IEEE Access*, *7*, 167260-167277, Article 8901118. https://doi.org/10.1109/ACCESS.2019.2953542

Li, A. W., & Bastos, G. S. (2020). Stock market forecasting using deep learning and technical analysis: A systematic review [Review]. *IEEE Access*, *8*, 185232-185242. https://doi.org/10.1109/ACCESS.2020.3030226

Li, B., & Xu, Z. (2021). A comprehensive bibliometric analysis of financial innovation. *Economic Research-Ekonomska Istraživanja*, 1-24.

Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., Min, H., & Deng, X. (2016). Empirical analysis: stock market prediction via extreme learning machine [Article]. *Neural Computing and Applications*, *27*(1), 67-78. https://doi.org/10.1007/s00521-014-1550-z

Liu, X., Zhang, J., & Guo, C. (2013). Full-text citation analysis: A new method to enhance scholarly networks. *Journal of the American Society for information Science and Technology*, *64*(9), 1852-1863.

Majhi, R., Panda, G., & Sahoo, G. (2009). Development and performance evaluation of FLANN based model for forecasting of stock markets [Article]. *Expert Systems with Applications*, *36*(3 PART 2), 6800-6808. https://doi.org/10.1016/j.eswa.2008.08.008

Majhi, R., Panda, G., Sahoo, G., Panda, A., & Choubey, A. (2008). Prediction of S&P 500 and DJIA stock indices using particle swarm optimization technique. 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence),

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, *13*(3), e0194889.

Markoulli, M. P., Lee, C. I., Byington, E., & Felps, W. A. (2017). Mapping Human Resource Management: Reviewing the field and charting future directions. *Human Resource Management Review*, *27*(3), 367-396.

Mazzocchi, F. (2018). Knowledge organization system (KOS): an introductory critical account. *KO Knowledge Organization*, *45*(1), 54-78.

Meho, L. I., & Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. *Journal of the American Society for information Science and Technology*, *59*(11), 1711-1726.

Mirmozaffari, M., Yazdani, M., Boskabadi, A., Ahady Dolatsara, H., Kabirifar, K., & Amiri Golilarz, N. (2020). A novel machine learning approach combined with optimization models for eco-efficiency evaluation. *Applied Sciences*, *10*(15), 5210.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213-228.

Moral Muñoz, J. A., Herrera Viedma, E., Santisteban Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review.

Nanda, S., Mahanty, B., & Tiwari, M. (2010). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*, *37*(12), 8793-8798.

Nandy, A., & Singh, P. K. (2020). Farm efficiency estimation using a hybrid approach of machine-learning and data envelopment analysis: Evidence from rural eastern India. *Journal of Cleaner Production*, *267*, 122106.

Nguyen, T.-T., & Yoon, S. (2019). A novel approach to short-term stock price movement prediction using transfer learning. *Applied Sciences*, *9*(22), 4745.

Nikolopoulos, C., & Fellrath, P. (1994). A hybrid expert system for investment advising. *Expert Systems*, *11*(4), 245-250.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions [Article]. *Artificial Intelligence Review*, *53*(4), 3007-3057. https://doi.org/10.1007/s10462-019-09754-z

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

Omidi, A., Nourani, E., & Jalili, M. (2011). Forecasting stock prices using financial data mining and Neural Network. 2011 3rd International Conference on Computer Research and Development,

Osinska, V., & Malak, P. (2016). Maps and mapping in scientometrics. *Methods and Tools for Studying Scientific Writing and its Users. Wrocław: Wroclaw University*, 59-73.

Oteng, D., Zuo, J., & Sharifi, E. (2021). A scientometric review of trends in solar photovoltaic waste management research. *Solar Energy*, *224*, 545-562.

Phua, P. K. H., Ming, D., & Lin, W. (2001). Neural network with genetically evolved algorithms for stocks prediction [Article]. *Asia-Pacific Journal of Operational Research*, *18*(1), 103-107.

Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers [Article]. *Applied Intelligence*, *26*(1), 25-33. https://doi.org/10.1007/s10489-006-0001-7

Qu, B., Leng, J., & Ma, J. (2019). Investigating the intensive redevelopment of urban central blocks using data envelopment analysis and deep learning: a case study of Nanjing, China. *IEEE Access*, *7*, 109884-109898.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for information Science and Technology*, *61*(9), 1871-1887.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, *29*(9), 2352-2449.

Rebai, S., Yahia, F. B., & Essid, H. (2020). A graphically based machine learning approach to predict secondary schools performance in Tunisia. *Socio-Economic Planning Sciences*, *70*, 100724.

Roh, T. H. (2007). Forecasting the volatility of stock price index. *Expert Systems with Applications*, *33*(4), 916-922.

Salehi, V., Veitch, B., & Musharraf, M. (2020). Measuring and improving adaptive capacity in resilient systems by means of an integrated DEA-Machine learning approach. *Applied Ergonomics*, *82*, 102975.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system [Article]. *ACM Transactions on Information Systems*, *27*(2), Article a12. https://doi.org/10.1145/1462198.1462204

Sharma, S., & Kaushik, B. (2018). Quantitative analysis of stock market prediction for accurate investment decisions in future [Review]. *Journal of Artificial Intelligence*, *11*(1), 48-54. https://doi.org/10.3923/jai.2018.48.54

Sheta, A. F., Ahmed, S. E. M., & Faris, H. (2015). A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. *Soft Computing*, *7*(8), 2.

Shrivas, A., & Sharma, S. K. (2018). A Robust Predictive Model for Stock Market Index Prediction using Data Mining Technique.

Singh, R., & Srivastava, S. (2017). Stock prediction using deep learning [Article]. *Multimedia Tools and Applications*, *76*(18), 18569-18584. https://doi.org/10.1007/s11042-016-4159-7

Somanathan, A. R., & Rama, S. K. (2020). A Bibliometric Review of Stock Market Prediction:: Perspective of Emerging Markets. *Applied Computer Systems*, *25*(2), 77-86.

Srinivasan, P., & Ibrahim, P. (2010). Forecasting stock market volatility of BSE-30 index using GARCH models. *Asia Pacific Business Review*, *6*(3), 47-60.

Su, C. H., Chen, T. L., Cheng, C. H., & Chen, Y. C. (2010). Forecasting the stock market with linguistic rules generated from the minimize entropy principle and the cumulative probability distribution approaches [Article]. *Entropy*, *12*(12), 2397-2417. https://doi.org/10.3390/e12122397

Su, H.-N., & Lee, P.-C. (2010). Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics*, *85*(1), 65-79.

Taherinezhad, A., & Alinezhad, A. (2022). Nations performance evaluation during SARS-CoV-2 outbreak handling via data envelopment analysis and machine learning methods. *International Journal of Systems Science: Operations & Logistics*, 1-18.

Tayal, A., Kose, U., Solanki, A., Nayyar, A., & Saucedo, J. A. M. (2020). Efficiency analysis for stochastic dynamic facility layout problem using meta‐heuristic, data envelopment analysis and machine learning. *Computational Intelligence*, *36*(1), 172-202.
Thaker, K., Charles, V., Pant, A., & Gherman, T. (2022). A DEA and random forest regression approach to studying bank efficiency and corporate governance. *Journal of the Operational Research Society*, *73*(6), 1258-1277.
Thakkar, A., & Chaudhari, K. (2021a). A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions [Review]. *Expert Systems with Applications*, *177*, Article 114800. https://doi.org/10.1016/j.eswa.2021.114800
Thakkar, A., & Chaudhari, K. (2021b). Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions [Article]. *Information Fusion*, *65*, 95-107. https://doi.org/10.1016/j.inffus.2020.08.019
Thavaneswaran, A., Appadoo, S. S., & Paseka, A. (2009). Weighted possibilistic moments of fuzzy numbers with applications to GARCH modeling and option pricing. *Mathematical and Computer Modelling*, *49*(1-2), 352-368.
Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting [Article]. *Expert Systems with Applications*, *40*(14), 5501-5506. https://doi.org/10.1016/j.eswa.2013.04.013
Tijssen, R. J., & Van Raan, A. F. (1994). Mapping changes in science and technology: Bibliometric co-occurrence analysis of the R&D literature. *Evaluation Review*, *18*(1), 98-115.
Ţiţan, A. G. (2015). The efficient market hypothesis: Review of specialized literature and empirical research. *Procedia Economics and Finance*, *32*, 442-449.
Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, *38*, 788-804.
Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches [Article]. *Decision Support Systems*, *50*(1), 258-269. https://doi.org/10.1016/j.dss.2010.08.028
Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523-538.
Van Eck, N. J., & Waltman, L. (2014). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of informetrics*, *8*(4), 802-823.
Van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, *111*(2), 1053-1070.
Van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for information Science and Technology*, *61*(12), 2405-2416.
Van Raan, A. (2019). Measuring science: basic principles and application of advanced bibliometrics. In *Springer handbook of science and technology indicators* (pp. 237-280). Springer.
Wang, M., Zhang, J., Jiao, S., & Zhang, T. (2019). Evaluating the impact of citations of articles based on knowledge flow patterns hidden in the citations. *PLoS ONE*, *14*(11), e0225276.
Wei, L.-Y., Chen, T.-L., & Ho, T.-H. (2011). A hybrid model based on adaptive-network-based fuzzy inference system to forecast Taiwan stock market. *Expert Systems with Applications*, *38*(11), 13625-13631.
Wei, Q., Zhang, J., & Zhang, X. (2000). An inverse DEA model for inputs/outputs estimate. *European Journal of Operational Research*, *121*(1), 151-163.
Wuni, I. Y., Shen, G. Q., & Osei-Kyei, R. (2019). Scientometric review of global research trends on green buildings in construction journals from 1992 to 2018. *Energy and buildings*, *190*, 69-85.
Yan, H., Wei, Q., & Hao, G. (2002). DEA models for resource reallocation and production input/output estimation. *European Journal of Operational Research*, *136*(1), 19-31.
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.

Yu, L., Chen, H., Wang, S., & Lai, K. K. (2009). Evolving least squares support vector machines for stock market trend mining [Article]. *IEEE Transactions on Evolutionary Computation*, *13*(1), 87-102. https://doi.org/10.1109/TEVC.2008.928176

Yudong, Z., & Lenan, W. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network [Article]. *Expert Systems with Applications*, *36*(5), 8849-8854. https://doi.org/10.1016/j.eswa.2008.11.028

Zhang, C.-T. (2013). A novel triangle mapping technique to study the h-index based citation distribution. *Scientific reports*, *3*(1), 1-5.

Zhang, J., Cui, S., Xu, Y., Li, Q., & Li, T. (2018). A novel data-driven stock price trend prediction system. *Expert Systems with Applications*, *97*, 60-69.

Zhang, L., Wang, F., Xu, B., Chi, W., Wang, Q., & Sun, T. (2018). Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI. *Neural Computing and Applications*, *30*(5), 1425-1444.

Zhao, Z.-Q., Zheng, P., Xu, S.-t., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, *30*(11), 3212-3232.

Zheng, X., Le, Y., Chan, A. P., Hu, Y., & Li, Y. (2016). Review of the application of social network analysis (SNA) in construction project management research. *International journal of project management*, *34*(7), 1214-1225.

Zhong, K., Wang, Y., Pei, J., Tang, S., & Han, Z. (2021). Super efficiency SBM-DEA and neural network for performance evaluation. *Information Processing & Management*, *58*(6), 102728.

Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*, *67*, 126-139.

Zhou, X., Pan, Z., Hu, G., Tang, S., & Zhao, C. (2018). Stock market prediction on high-frequency data using generative adversarial nets. *Mathematical Problems in Engineering*, *2018*.

Zhu, J. (2020). DEA under big data: Data enabled analytics and network data envelopment analysis. *Annals of operations research*, 1-23.

Zhu, N., Zhu, C., & Emrouznejad, A. (2021). A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies [Article]. *Journal of Management Science and Engineering*, *6*(4), 435-448. https://doi.org/10.1016/j.jmse.2020.10.001