GMMC Journal of Interdisciplinary Studies
Vol. 13, December 2024, pp. 116-141
ISSN : 2392-4519 (Print), 3021-9086 (Online)
Journal homepage: http://gmmcjournal.edu.np

Bibliometric Analysis of Machine Learning Models for Stock Price Prediction

***Prajjwol Bikram Khadka[1], Dipendra Karki, PhD[2]**
[1]Faculty, Kathmandu Don Bosco College, Nepal
ORCID ID: https://orcid.org/0009-0006-0541-9582
[2]Faculty of Management, Tribhuvan University, Nepal
ORCID ID: https://orcid.org/0000-0001-9045-7423
**Corresponding Author's Email: prajjwolkhadka2@gmail.com*

Received: August, 2024
Revised: September, 2024
Accepted: November, 2024
Available online: December, 2024
DOI: https://doi.org/10.3126/jis.v13i1.73347

ABSTRACT

Machine learning models are becoming more popular for predicting stock prices. Their advanced algorithms often perform better than traditional statistical methods. This study aims to carry out a bibliometric analysis using VosViewer. It focuses on the number, variety, and connections of studies related to stock price prediction using Support Vector Machines (SVM), Random Forests, and Deep Learning techniques. Boolean operators were used to search for and filter relevant articles from the Dimensions database. We conducted a manual review to select only high-quality studies, resulting in 55 peer-reviewed articles for analysis.

The analysis highlights important trends, influential studies, and collaborative networks in the field. This provides a clear overview of how machine learning models are developing for stock price prediction. There has been a notable rise in publications since 2013, with major contributions from institutions in the United States, the United Kingdom, India, and China. Emerging themes, especially deep learning applications, have been identified. However, there is still a lack of in-depth academic evaluation of these models.

This study stresses the need for more empirical research to address algorithmic biases and improve model reliability. The findings offer useful insights for researchers and practitioners, guiding future studies and helping to develop more accurate stock price prediction models.

Keywords: Bibliometric analysis, deep learning, machine learning, stock price prediction

INTRODUCTION

Stock price prediction has always been a popular as well as significant topic of research in finance due to its potential for high economic returns. Accurate predictions is highly beneficial for stakeholders like academic researchers and financial practitioners as it yields substantial profits (Anh & Son, 2024). The traditional way of predicting stock price often relies on statistical models and economic theories. However, in the recent times machine learning models are being used as powerful tools for predicting stock prices in the presence of big data and the increasing availability of financial data (Breiman, 2001).

There are a number of ways that machine learning models are better than traditional statistical methods. By learning from historical data, they can handle large datasets, find sophisticated patterns, and make predictions that are more accurate. Support Vector Machines (SVM), Random Forests, and Deep Learning are some of the machine learning models that have shown a lot of potential. Each model has its own strengths. For instance, SVMs are known for being stable in high-dimensional spaces (Vapnik, 2000); Random Forests can handle large datasets and reduce overfitting (Breiman, 2001); and Deep Learning models can learn complex patterns through multiple layers of abstraction (Hochreiter & Schmidhuber, 1997a).

It is a well-known fact that stock price prediction has been considered one of the most important and significant research domains in the field of financial markets, which involves making forecasts on stock prices based on historical data and various predictive models. In the methods used in the traditional approaches of stock price prediction, various statistical models and economic theories were used: Among these, the Autoregressive Integrated Moving Average (ARIMA) model was widely used due to its simplicity and accuracy for modeling the linear relationship in time series data (Wilson, 2016). The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, which is another traditional method, concentrated its attention to the analytics of volatility of stock prices, which is extremely significant for managing risks (Bollerslev, 1986).

However, the traditional approaches mainly do not address many underlying, non-linear relationships often present in financial markets. Due to this limitation, more sophisticated and modern tools like machine learning models are being adopted to precisely predict stock prices. These methods leverage the power of machine learning and artificial intelligence to enhance prediction accuracy by uncovering hidden patterns in the data that traditional models might miss (Zhang, 2003).

MATERIALS AND TECHNIQUES

Review of Traditional and Contemporary Methods of Stock Price Prediction

Overview of Machine Learning Models

Support Vector Machines (SVM)

Support Vector Machines (SVM) are types of models belong to the supervised learning class and can be employed in classification and regression analysis (Vapnik, 2000). Their workings are based on the ability to identify the hyperplane that defines the maximum margin between different classes of data points. The application of SVMs is ideal for high dimensional space and they are not very much affected by over fitting issues hence can be used in stock price prediction. SVMs have been used to predict stock price movements and volatility whereby some studies have revealed that the models outperform traditional methods in certain scenarios (Huang et al., 2005; Kim, 2003)

Other similar researches have also confirmed the use of Support Vector Machines in stock price prediction. For instance, a study conducted on the comparative analysis of SVM with other machine learning techniques and observed that it performs very well in High Frequency trading environment (Patel et al., 2015). Moreover, SVM has been combined with other techniques, such as genetic algorithms, to further enhance predictive performance (Al-Ashaab et al., 2012).

Random Forests Techniques

Proposed by Breiman (2001), random forests are complex learning techniques which generate one decision tree when the model is trained and then the mode of the classes (in classification problem) or the mean forecast (in regression problem) of the trees is returned. This minimizes overfitting and enhances the accuracy of the model since and fosters averaging of the resultant trees. Random forests have been implemented more often in predicting the stock prices due to their ability to handle large datasets, robustness to noise, and capacity to model complex interactions between features (Almeida & Figueiredo, 2010; Blocher, 2016).

Further research in the matter deems the random forest method highly stable and readily applicable to financial prediction. For instance, a study established that random forests outperformed other branches in stock market forecasting in various market environments (Ballings et al., 2015). Moreover, in the context of the regression analysis, (Deloof and Vanacker (2018) apply random forests for the assessment of stock price crashes, thus proving the model's applicability to extreme outcomes.

Deep Learning Techniques

Among the modern approaches in predictive modeling, the application of deep learning, and in particular, neural networks with multiple layers (Deep Neural Networks) can be referred to as groundbreaking. CNNs and RNNs and its variant such as LSTM are believed to be effective for stock price prediction. CNNs can keep the spatial hierarchies in the data features while LSTMs are good in time series forecasting because of the long-term dependency (Hochreiter & Schmidhuber, 1997b; LeCun et al., 2015). The analyses of the stock price prediction using deep learning models have shown that deep learning techniques are more efficient than the traditional and other machine methods (Eskandari-Khanghahi et al., 2018; Xie et al., 2017).

The significance of the role of deep learning in the financial markets have further solidified with the recent advancements in machine learning. For instance, Nelson et al. (2017) performed a study to predict the movements of the stock prices with the help of LSTMs and noted that the implementation of deep LSTMs provided significantly better results as compared to the conventional models. Similarly, Heaton et al. (2016)studied usage of deep reinforcement learning for trading strategies that belongs to the deep learning with decision-making processes.

METHODOLOGY

Data Collection and Search Strategy

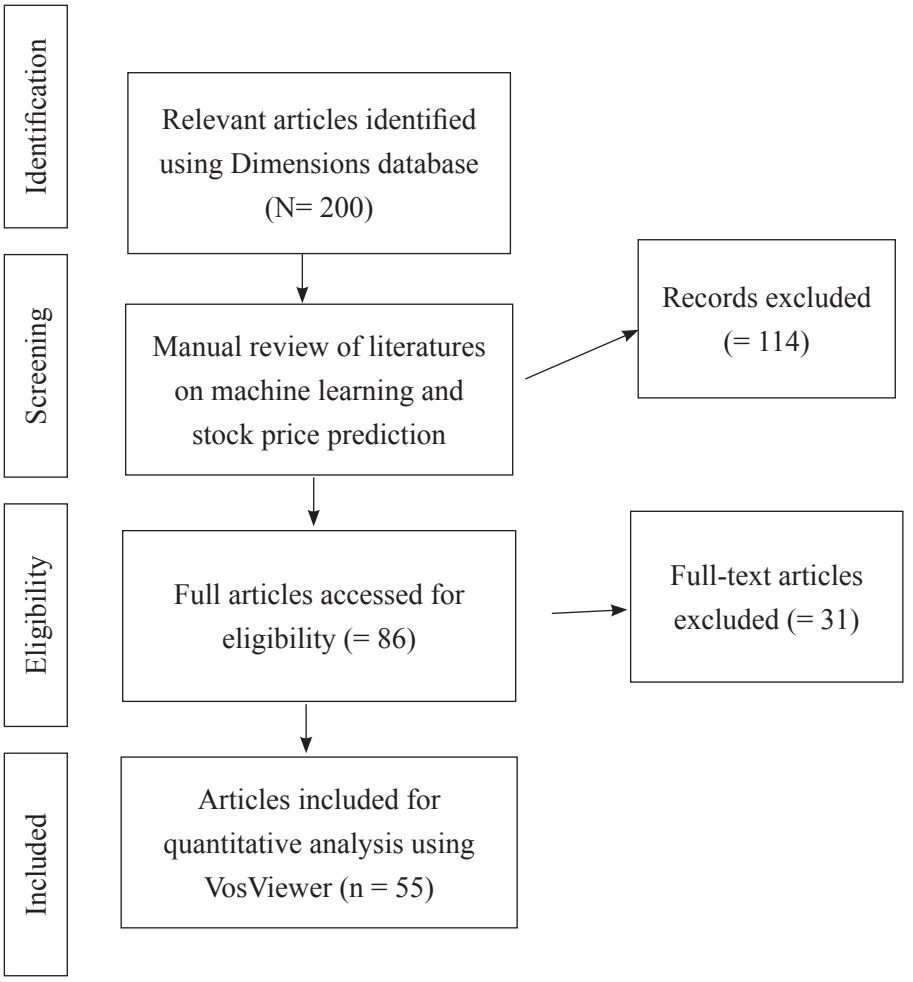
The data required for this bibliometric study were obtained from the Dimensions database, which is a popular and a large database containing data of academic publications and citation data (Thelwall, 2018). The search queries were built with the help of Boolean operators so that the obtained data set was relevant to the given topic. The specific search query used was:

("Stock Price Prediction" OR "Stock Market Prediction" OR "Financial Market Prediction") AND ("Machine Learning" OR "ML") AND ("Support Vector Machine" OR "SVM") AND ("Random Forest" OR "RF") AND ("Deep Learning" OR "Neural Networks")

There was negligible research in the domain of stock price prediction using machine learning models before 2013. Therefore, the search was limited to articles from 2013 to 2024 to capture recent trends and advancements in the field. An initial sample of 200 journal articles was identified. Out of these 200 articles, a process of relevance appraisal and quality check was done by the researchers, thus only 55 relevant and quality articles were used in this study.

Figure 1

Search strategy to select relevant studies for analysis recommended by PRISMA framework



Source: (Rethlefsen et al., 2021)

Bibliometric Tools

The bibliometric software, VosViewer was used to generate visualizations that illustrate the relationships and trends within the selected body of literature (Van Eck & Waltman, 2010).

RESULTS

Based on a thorough bibliometric analysis, significant trends emerged in the study of machine learning models for predicting stock prices. Based on our analysis of 55 selected articles published between 2013 and 2024, there has been a clear upward trend in the number of publications over the years. The trend observed here suggests a rising interest and progress in the application of machine learning techniques to predict financial market price movements.

Co-Authorship Analysis

Co-authorship analysis was applied in order to identify collaboration patterns among authors and institutions. Thus, analyzing the co-authorship networks, it is possible to identify key researchers and research groups actively working in the domain of machine learning for stock price prediction (Newman, 2004).

Co-Authorship- Authors

The following table shows the top ten notable contributors to machine learning models for stock price prediction. Each author's document count, citations, and link strength are listed. Nelson Areal, Paulo Cortez, and Nuno Oliveira are highly respected authors with 262 citations and 2 links per work, making them prominent in this field. Similarly, Bruno Miranda Henrique, Herbert Kimura, and Vinicius Amorim Sobreiro have 197 citations, showing their strong scientific influence. Ruizhu Han, Yilin Ma, and Weizhong Wang have an impressive 145 citations each for their combined work. Ash Booth, with 113 citations, also plays a key role in the collaborative network.

These authors form a strong network, emphasizing the importance of collaboration and knowledge sharing in machine learning research for stock price prediction. Their interdependence and collective efforts show how working together is vital to achieving excellence in this field.

Table 1

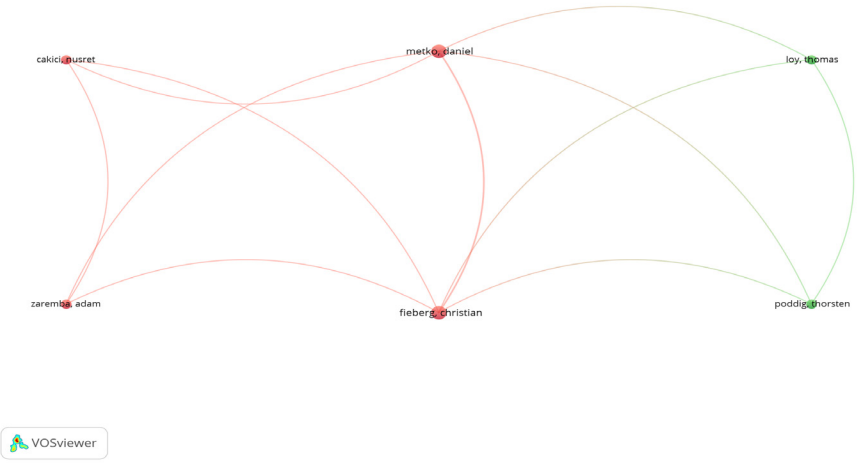
Co-Authorship- Authors

ID	Author	Documents	Citations	Total Link Strength
8	areal, nelson	1	262	2
23	cortez, paulo	1	262	2
102	oliveira, nuno	1	262	2
56	henrique, bruno miranda	1	197	2
76	kimura, herbert	1	197	2
137	sobreiro, vinicius amorim	1	197	2

ID	Author	Documents	Citations	Total Link Strength
53	han, ruizhu	1	145	2
86	ma, yilin	1	145	2
147	wang, weizhong	1	145	2
13	booth, ash	1	113	2

Figure 2

Bibliometric Map of Co-Authorship Using Author Names



Co-Authorship- Organizations

The following is a list of the top ten important organizations involved in researching machine learning models for predicting stock prices. The University of Minho and the University of Brasília have gained significant attention, with 262 and 197 citations respectively, showing their strong influence. Other notable contributors include Southeast University and the University of Southampton. The Catholic University of Lyon and EDC Paris Business School have a strong collaborative relationship, as shown by their link strengths of 2. Similarly, The Federal University of Paraná and Pontifícia Universidade Católica do Paraná have made notable contributions, with a total link strength of 3, highlighting their impactful collaboration.

The co-authorship network visualization shows the connections between different institutions, emphasizing strong international research collaborations. Notable organizations like Montpellier Business School and Concordia University also stand out in this network.

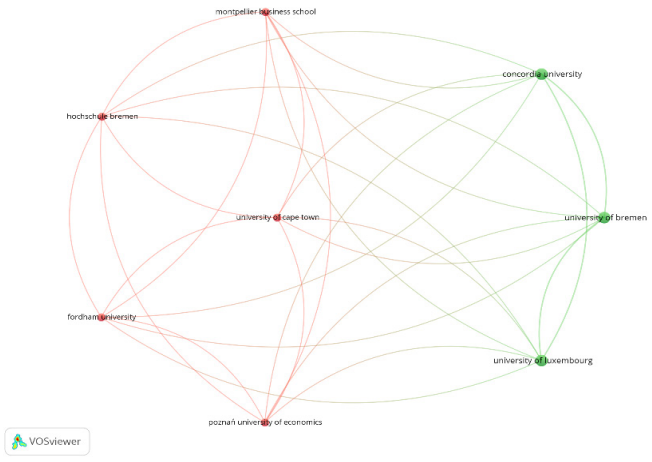
Table 2

Co-Authorship- Organizations

ID	Organization	Documents	Citations	Total Link Strength
91	university of minho	1	262	0
84	university of brasília	1	197	0
71	southeast university	1	145	0
98	university of southampton	1	113	0
11	catholic university of lyon	1	67	2
20	edc paris business school, observatory and research center on entrepreneurship (ocre), department of entrepreneurship and digital transformation, 70 galerie des damiers - paris la défense 1, 92415 courbevoie cedex, france. electronic address: wbenarfi@edcparis.edu.	1	67	2
96	university of sfax	1	67	2
25	federal university of paraná	1	64	3
58	pontifícia universidade católica do paraná	1	64	3
72	southwestern university of finance and economics	1	64	1

Figure 3

Bibliometric Map of Co-Authorship Organizations Using Organization Names



Co-Authorship- Countries

The table highlights the valuable research conducted by different countries on machine learning models for stock price prediction. China leads with nine documents and 282 citations, showing a strong research output and influence. Brazil and Portugal also have impressive citation counts, with 267 and 262 respectively. The United Kingdom has the highest total link strength of 5, indicating extensive international collaboration. It has contributed five documents and received 199 citations, showing its importance in this field. France and the United States have demonstrated strong collaboration, resulting in a total link strength of 7.

Countries like Tunisia, India, Greece, and South Korea have made notable contributions, showcasing a globally connected and diverse research community. The visualization of the co-authorship network highlights these global collaborations, emphasizing the interdependence among nations and their collective efforts to advance research in this area.

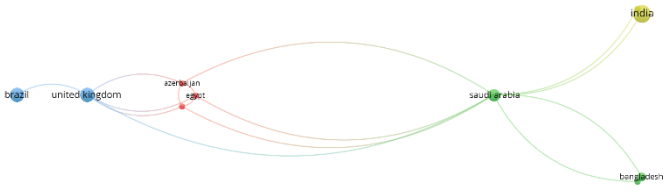
Table 3

Co-Authorship- Countries

ID	Country	Documents	Citations	Total Link Strength
5	China	9	282	2
3	Brazil	5	267	1
19	Portugal	1	262	0
29	United Kingdom	5	199	5
8	France	2	74	7
30	United States	5	69	7
27	Tunisia	1	67	1
11	India	8	35	2
10	Greece	3	32	0
24	South Korea	1	31	1

Figure 4

Bibliometric Map of Co-Authorship Countries Using Country Names



Citation Analysis

Citation analysis deals with counting the number of times a paper is cited by other papers. It proved useful for measuring the role and activity of articles, authors, and journals in the given field (Garfield, 1972).

Citation Analysis- Authors

The table shows a citation analysis of authors researching machine learning models for stock price prediction. Nelson Areal, Paulo Cortez, and Nuno Oliveira are recognized for their influential work, with 262 citations each and a total link strength of 12. This shows their strong impact and the solid collaborative networks they have built. Similarly, Bruno Miranda Henrique, Herbert Kimura, and Vinicius Amorim Sobreiro have made significant contributions, shown by their 197 citations each. Ruizhu Han, Yilin Ma, and Weizhong Wang also have high influence, with 145 citations each and a link strength of 13, reflecting their impact.

Ash Booth stands out with a strong number of citations and a solid link strength, making him a key player in this field. The network visualization shows the connections and impact of these authors, highlighting their collaborative efforts and important contributions to this research area.

Table 4

Citation Analysis- Authors

ID	Author	Documents	Citations	Total Link Strength
8	areal, nelson	1	262	12
23	cortez, paulo	1	262	12
102	oliveira, nuno	1	262	12
56	henrique, bruno miranda	1	197	3
76	kimura, herbert	1	197	3
137	sobreiro, vinicius amorim	1	197	3
53	han, ruizhu	1	145	13
86	ma, yilin	1	145	13
147	wang, weizhong	1	145	13
13	booth, ash	1	113	3

Figure 5

Bibliometric Map of Citation Analysis-Authors Using Author Names

Citation Analysis- Organizations

The table shows a citation analysis of organizations that have contributed to research on machine learning models for stock price prediction. The University of Minho and the University of Brasília stand out for their significant contributions in terms of citations and impact. Similarly, Southeast University has an impressive 145 citations and the highest total link strength of 5, reflecting its strong collaborative efforts.

Several other organizations, including the University of Southampton, the Catholic University of Lyon, and EDC Paris Business School, have made important contributions with varying citation counts and link strengths. The network visualization highlights the University of Hamburg's strong connections with institutions such as Brunel University London and Swansea University. This demonstrates a wide-reaching collaborative network and significant citation impact in this research area.

Table 5

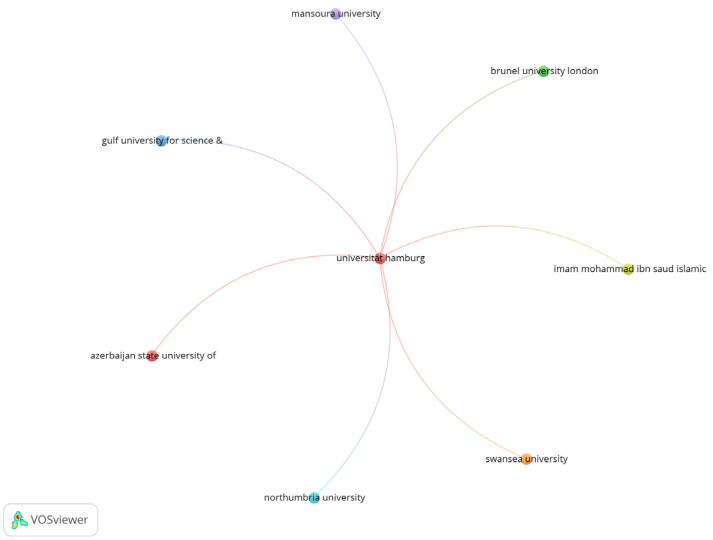
Citation Analysis- Organizations

ID	Organization	Documents	Citations	Total Link Strength
91	university of minho	1	262	2
84	university of brasília	1	197	1
71	southeast university	1	145	5
98	university of southampton	1	113	1
11	catholic university of lyon	1	67	0

ID	Organization	Documents	Citations	Total Link Strength
20	edc paris business school, observatory and research center on entrepreneurship (ocre), department of entrepreneurship and digital transformation, 70 galerie des damiers - paris la défense 1, 92415 courbevoie cedex, france. electronic address: wbenarfi@edcparis.edu.	1	67	0
96	university of sfax	1	67	0
25	federal university of paraná	1	64	0
58	pontifícia universidade católica do paraná	1	64	0
72	southwestern university of finance and economics	1	64	1

Figure 6

Bibliometric Map of Citation Analysis-Organizations Using Organization Names



Citation Analysis- Countries

The table shows a citation analysis of countries involved in research on machine learning models for stock price prediction. China leads with nine documents and 282 citations, showing strong research output and a total link strength of 4. Brazil has five documents and 267 citations, making it a key contributor. Portugal has one document with 262 citations and a link strength

of 2. The United Kingdom has also made significant contributions, with five documents and 199 citations, and a link strength of 3.

France and the United States, along with other countries, have made moderate contributions, reflected in lower citation counts. Tunisia, India, Greece, and South Korea, despite having fewer citations, still make a notable impact on the research landscape. The network visualization highlights strong collaborative connections among these countries, showing a global effort to advance this field.

Table 6

Citation Analysis- Countries

ID	Country	Documents	Citations	Total Link Strength
5	China	9	282	4
3	Brazil	5	267	1
19	Portugal	1	262	2
29	United Kingdom	5	199	3
8	France	2	74	0
30	United States	5	69	0
27	Tunisia	1	67	0
11	India	8	35	2
10	Greece	3	32	2
24	South Korea	1	31	0

Figure 7

Bibliometric Map of Citation Analysis-Organizations Using Country Names

Co-Citation Analysis

Co-citation analysis deals with analysing the frequency with which pairs of documents are cited together by subsequent papers. The co-citation analysis is important in defining the seminal works and the pattern of the development of research on machine learning for stock price prediction (Small, 1973b).

Co-Citation Analysis- Cited Authors

The table shows the co-citation analysis of key authors in the study of machine learning models for predicting stock prices. Eugene F. Fama stands out as a highly influential figure with 36 citations and a total link strength of 3337. Similarly, Kenneth R. French and Guofu Zhou each have 19 citations and strong link strengths of 2178 and 2428, respectively. Notable contributors like David Enke, Bryan Kelly, and Christopher Krauss have made significant contributions with 16 citations and strong links.

Rangan Gupta, Omer Berat Sezer, and Dacheng Xiu each have 14 citations, showing their influence through co-citation connections. Tim Bollerslev rounds out the list with 13 citations and a link strength of 1044. The network visualization highlights the complex connections among these authors, emphasizing the strong and influential co-citation relationships shaping this research field.

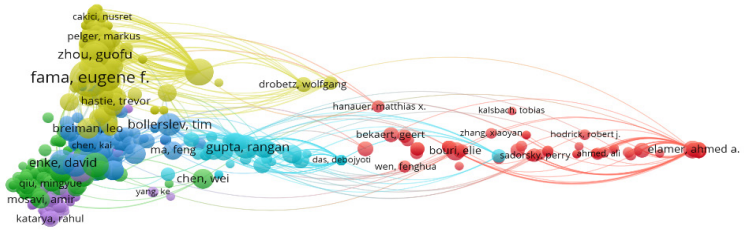
Table 7

Co-Citation Analysis- Cited Authors

ID	Author	Citations	Total Link Strength
1118	Fama, Eugene F.	36	3337
1203	French, Kenneth R.	19	2178
4814	Zhou, Guofu	19	2428
1089	Enke, David	16	1051
2027	Kelly, Bryan	16	1596
2171	Krauss, Christopher	16	1280
1449	Gupta, Rangan	14	1104
3640	Sezer, Omer Berat	14	1054
4464	Xiu, Dacheng	14	1231
438	Bollerslev, Tim	13	1044

Figure 8

Bibliometric Map of Co-Citation Analysis Using Cited Author Names



Co-Citation Analysis- Cited Sources

The table shows the co-citation analysis of key sources cited in research on machine learning models for stock price prediction. "Expert Systems with Applications" leads with 270 citations and a total link strength of 12,997, showing its major influence. "The Journal of Finance" and "Journal of Financial Economics" follow with 91 and 86 citations, and strong link strengths of 5907 and 5684, respectively.

Other important sources include "Applied Soft Computing" with 75 citations, "SSRN Electronic Journal" with 63 citations, and "Review of Financial Studies" with 58 citations. These sources demonstrate a substantial impact and strong co-citation relationships. "IEEE Access," "Neurocomputing," "Energy Economics," and "International Journal of Forecasting" also show notable influence with significant citations and link strengths.

The network visualization shows the connections among these sources, highlighting the dense co-citation relationships that shape the research landscape in this field.

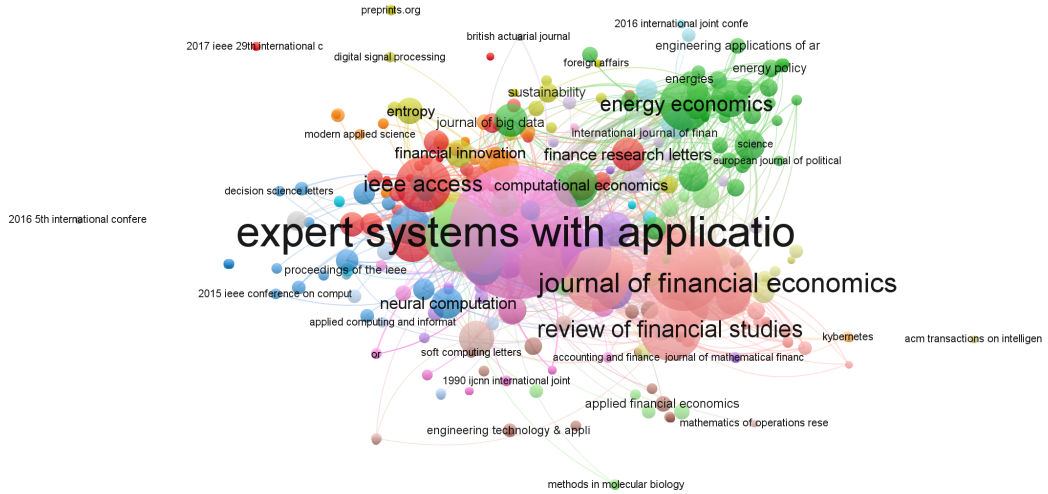
Table 8

Co-Citation Analysis- Cited Sources

ID	Source	Citations	Total Link Strength
256	expert systems with applications	270	12997
686	the journal of finance	91	5907
454	journal of financial economics	86	5684
140	applied soft computing	75	4007
661	ssrn electronic journal	63	4378
631	review of financial studies	58	3999
288	ieee access	48	2047
554	neurocomputing	46	2566
234	energy economics	42	2864
361	international journal of forecasting	40	2339

Figure 9

Bibliometric Map of Co-Citation Analysis Using Cited Sources



Co-Citation Analysis- Cited References

The table shows the co-citation analysis of key sources cited in research on machine learning models for stock price prediction. “Expert Systems with Applications” leads with 270 citations and a total link strength of 12,997, showing its major influence. “The Journal of Finance” and “Journal of Financial Economics” follow with 91 and 86 citations, and strong link strengths of 5907 and 5684, respectively.

Other important sources include “Applied Soft Computing” with 75 citations, “SSRN Electronic Journal” with 63 citations, and “Review of Financial Studies” with 58 citations. These sources demonstrate a substantial impact and strong co-citation relationships. “IEEE Access,” “Neurocomputing,” “Energy Economics,” and “International Journal of Forecasting” also show notable influence with significant citations and link strengths.

The network visualization shows the connections among these sources, highlighting the dense co-citation relationships that shape the research landscape in this field.

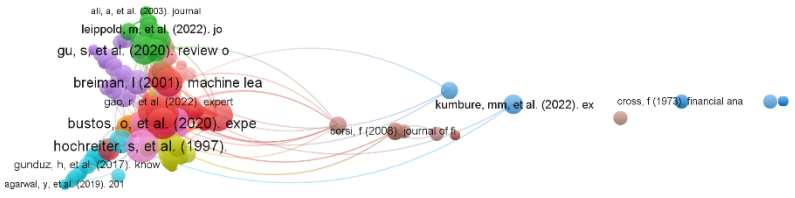
Table 9

Co-Citation Analysis- Cited References

ID	Cited Reference	Citations	Total Link Strength
869	hochreiter, s, et al. (1997). neural computation, 9(8), 1735-1780	11	373
289	bustos, o, et al. (2020). expert systems with applications, 156113464	10	285
401	chong, e, et al. (2017). expert systems with applications, 83187-205	10	371
657	fischer, t, et al. (2018). european journal of operational research, 270(2), 654-669	10	411
445	cortes, c, et al. (1995). machine learning, 20(3), 273-297	9	149
265	breiman, l (2001). machine learning, 45(1), 5-32	8	286
611	fama, ef (1970). the journal of finance, 25(2), 383	8	229
762	gu, s, et al. (2020). review of financial studies, 33(5), 2223-2273	8	403
1026	kara, y, et al. (2011). expert systems with applications, 38(5), 5311-5319	7	141
1250	long, w, et al. (2019). knowledge-based systems, 164163-173	7	320

Figure 10

Bibliometric Map of Co-Citation Analysis Using Cited References



Bibliographic Coupling

Under bibliographic coupling, the coupling of two documents is defined by the fact that one of the documents has cited the other or both have cited one or several other documents. The bibliographic coupling reveals the extent of connection between the research papers through the references given which demonstrate new research associations (Kessler, 1963).

Bibliographic Coupling- Authors

The table shows the bibliographic coupling of key authors in the study of machine learning models for predicting stock prices. Nelson Areal, Paulo Cortez, and Nuno Oliveira collaborated on a highly recognized document, receiving 262 citations and a total link strength of 178. This indicates strong connections in their research work. Bruno Miranda Henrique, Herbert Kimura, and Vinicius Amorim Sobreiro have achieved 197 citations, reflecting strong academic ties and a link strength of 196.

The work by Ruizhu Han, Yilin Ma, and Weizhong Wang shows a strong connection, as seen by their high number of citations and link strength. This suggests they share common research interests and often reference each other's work. Ash Booth, with 113 citations and a link strength of 247, also shows strong bibliographic coupling.

The network visualization highlights the complex connections between these authors, showing their collaborative spirit and the shared references that connect their research.

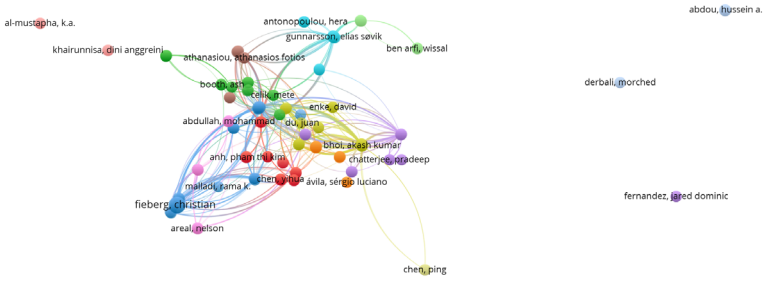
Table 10

Bibliographic Coupling- Authors

ID	Author	Documents	Citations	Total Link Strength
8	Areal, Nelson	1	262	178
23	Cortez, Paulo	1	262	178
102	Oliveira, Nuno	1	262	178
56	Henrique, Bruno Miranda	1	197	196
76	Kimura, Herbert	1	197	196
137	Sobreiro, Vinicius Amorim	1	197	196
53	Han, Ruizhu	1	145	416
86	Ma, Yilin	1	145	416
147	Wang, Weizhong	1	145	416
13	Booth, Ash	1	113	247

Figure 11

Bibliometric Map of Bibliographic Coupling Using Author Names



Bibliographic Coupling- Organizations

The table shows the bibliographic coupling of key organizations involved in research on machine learning models for stock price prediction. One document from the University of Minho has been cited 262 times, showing its strong impact in this field. It also has a total link strength of 45, highlighting its influence. The data shows that the University of Brasília has 197 citations and a link strength of 85, while Southeast University has 145 citations and a high link strength of 182.

Several prestigious institutions, such as the University of Southampton, the Catholic University of Lyon, EDC Paris Business School, and the University of Sfax, demonstrate strong bibliographic coupling. This suggests a significant overlap in their research references. The Federal University of Paraná and Pontifícia Universidade Católica do Paraná both have a strong link strength of 202, indicating a high level of interconnectedness.

The network visualization shows the complex connections between these institutions, emphasizing the collaborative and interconnected nature of their research.

Table 11

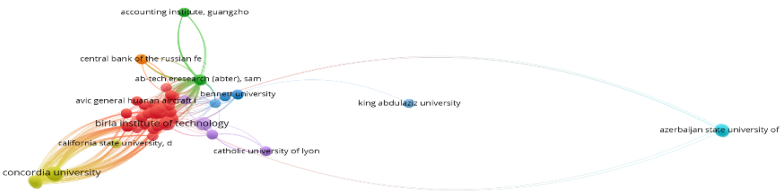
Bibliographic Coupling- Organizations

ID	Organization	Documents	Citations	Total Link Strength
91	University Of Minho	1	262	45
84	University Of Brasília	1	197	85
71	Southeast University	1	145	182
98	University Of Southampton	1	113	105

ID	Organization	Documents	Citations	Total Link Strength
11	Catholic University Of Lyon	1	67	196
20	Edc Paris Business School, Observatory And Research Center On Entrepreneurship (Ocre), Department Of Entrepreneurship And Digital Transformation, 70 Galerie Des Damiers - Paris La Défense 1, 92415 Courbevoie Cedex, France. Electronic Address: Wbenarfi@ Edcparis.Edu.	1	67	196
96	University of Sfax	1	67	196
25	Federal University of Paraná	1	64	202
58	Pontifícia Universidade Católica Do Paraná	1	64	202
72	Southwestern University of Finance And Economics	1	64	155

Figure 12

Bibliometric Map of Bibliographic Coupling Using Organization Names



Bibliographic Coupling- Countries

The table shows the bibliographic coupling of countries involved in studying machine learning models for predicting stock prices. China leads with nine documents, 282 citations, and a total link strength of 631, highlighting its strong position in the research field. Brazil has five documents, 267 citations, and a link strength of 249, while Portugal has one document with 262 citations and a link strength of 31. The United Kingdom has five documents and 199 citations, with a strong link strength of 724, showing a high level of bibliographic connections.

The United States and France also stand out with link strengths of 869 and 859, indicating strong interconnectedness. India, with eight documents and a link strength of 604,

has a notable impact as well. The network visualization shows the complex bibliographic connections between these countries, emphasizing the global reach and collaborative efforts in this research area.

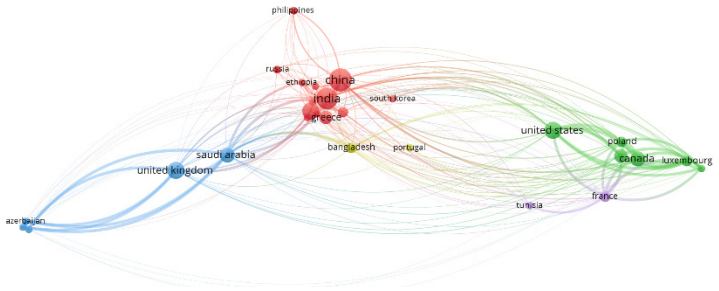
Table 12

Bibliographic Coupling- Countries

ID	Country	Documents	Citations	Total Link Strength
5	china	9	282	631
3	brazil	5	267	249
19	portugal	1	262	31
29	united kingdom	5	199	724
8	france	2	74	859
30	united states	5	69	869
27	tunisia	1	67	102
11	india	8	35	604
10	greece	3	32	68
24	south korea	1	31	70

Figure 13

Bibliometric Map of Bibliographic Coupling Using Country Names



Summary of Results

Our analysis reveals several key findings about the research landscape of machine learning models for stock price prediction. There has been a notable increase in publications over the years, showing strong interest and continuous progress in this field. The co-authorship analysis highlights the importance of collaboration among researchers, with certain authors and institutions emerging as central figures in the network.

Co-citation analysis has identified major contributions in areas like SVM, Random

Forests, and Deep Learning techniques. These works have laid a strong foundation for further research. Bibliographic coupling has helped pinpoint distinct research clusters focused on different themes, such as SVM applications, Random Forests, Deep Learning for time series, and hybrid models.

To build on this study, we present meta tables that summarize key insights from the bibliometric analysis.

Table 13

Key Machine Learning Techniques for Stock Price Prediction

Technique	Strengths	Weaknesses	Notable Applications/ Studies
Support Vector Machines (SVM)	Stable in high-dimensional spaces, handles classification and regression well	Can be computationally intensive with large datasets	Kim (2003), Huang et al. (2005)
Random Forests	Reduces overfitting, robust to noise, good for large datasets	Prone to becoming complex and difficult to interpret	Ballings et al. (2015), Blocher (2016)
Deep Learning (CNN, RNN, LSTM)	Can capture complex, non-linear relationships in time series data	Requires large datasets and is resource-intensive	Nelson et al. (2017), Hochreiter & Schmidhuber (1997)

Table 14

Top Contributing Institutions in Machine Learning for Stock Prediction

Institution	Number of Publications	Total Citations	Notable Research Contributions
University of Minho	1	262	Pioneering research on machine learning models for stock prediction
University of Brasília	1	197	Significant studies on SVM and Random Forest applications
Southeast University	1	145	Research contributions on deep learning models

Table 15

Country-wise Contribution in Research (2013-2024)

Country	Number of Publications	Total Citations	Total Link Strength	Notable Collaborations
China	9	282	2	Strong collaboration with Brazil
Brazil	5	267	1	Notable contributions to Random Forest techniques
United Kingdom	5	199	5	Extensive international collaboration

The above meta table information offers a comprehensive overview of key research findings, trends, and patterns.

DISCUSSION

This study explored and analyzed the research landscape of machine learning models used for stock price prediction through a bibliometric approach. The main goal was to identify trends, key contributors, and influential works to understand the field's evolution and current state.

The results show a significant rise in research activity over the past decade. This increase highlights a growing interest in applying machine learning to financial market predictions. Key methods like Support Vector Machines (SVM), Random Forests, and Deep Learning models have been widely studied. This reflects the demand for more accurate and complex predictive tools in the industry.

Our analysis of co-authorship networks shows the collaborative nature of this field. Prominent researchers and institutions have made major contributions, driving advancements in machine learning applications for stock price prediction. Understanding these collaborations offers insights into key players and potential areas for future collaboration.

Through co-citation analysis, we identified foundational works, such as Breiman's (2001) Random Forests model and Hochreiter and Schmidhuber's (1997) LSTM networks. These works are frequently referenced and have shaped predictive modelling advancements.

Bibliographic coupling revealed distinct research clusters. These clusters focus on various methods, including SVM applications for classification, Random Forests for complex data interactions, and deep learning models for time series predictions. This shows the diverse

approaches researchers take to handle the complexities and non-linear behaviors of financial market data.

Hence, our findings align with the study's objectives by providing a comprehensive view of research progress in applying machine learning to stock price prediction. Future research should continue exploring new techniques, addressing algorithmic biases, and promoting more collaboration across institutions and countries to improve predictive accuracy and model reliability.

CONCLUSION

This study conducted a detailed bibliometric analysis of machine learning models used for stock price prediction. The focus was on Support Vector Machines (SVM), Random Forests, and Deep Learning techniques. Our analysis revealed several important trends and insights. There has been a noticeable increase in publications in recent years, reflecting a growing interest and progress in using machine learning for predicting stock prices. By analyzing co-authorship, we identified key researchers, institutions, and countries leading this field. This analysis highlights the importance of collaborative networks in driving research progress. Additionally, co-citation analysis uncovered influential papers that have significantly impacted later research. Breiman's work on Random Forests (2001) and Hochreiter and Schmidhuber's work on LSTM networks (1997) stand out as highly influential. Through bibliographic coupling, we identified distinct research clusters focused on various machine learning techniques. These include SVM applications, Random Forests, and Deep Learning for time series prediction. Citation analysis further highlighted the significance of certain articles, as those with high citation counts indicate strong influence and recognition among scholars.

REFERENCES

Al-Ashaab, A., Molyneaux, M., Doultsinou, A., Brunner, B., Martínez, E., Moliner, F., Santamaría, V., Tanjore, D., Ewers, P., & Knight, G. (2012). Knowledge-based environment to support product design validation. *Knowledge-Based Systems*, *26*, 48–60. https://doi.org/10.1016/j.knosys.2011.06.019

Almeida, A. de, & Figueiredo, M. B. (2010). A particular approach for the Three-dimensional Packing Problem with additional constraints. *Computers & Operations Research*, *37*(11), 1968–1976. https://doi.org/10.1016/j.cor.2010.01.010

Anh, N. Q., & Son, H. X. (2024). *Transforming Stock Price Forecasting: Deep Learning*

Architectures and Strategic Feature Engineering (pp. 237–250). https://doi.org/10.1007/978-3-031-68208-7_20

Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, *42*(20), 7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013

Blocher, J. (2016). Network externalities in mutual funds. *Journal of Financial Markets*, *30*, 1–26. https://doi.org/10.1016/j.finmar.2016.04.001

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Deloof, M., & Vanacker, T. (2018). The recent financial crisis, start□up financing and survival. *Journal of Business Finance & Accounting*, *45*(7–8), 928–951. https://doi.org/10.1111/jbfa.12319

Eskandari-Khanghahi, M., Tavakkoli-Moghaddam, R., Taleizadeh, A. A., & Amin, S. H. (2018). Designing and optimizing a sustainable supply chain network for a blood platelet bank under uncertainty. *Engineering Applications of Artificial Intelligence*, *71*, 236–250. https://doi.org/10.1016/j.engappai.2018.03.004

Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). *Deep Learning in Finance*.

Hochreiter, S., & Schmidhuber, J. (1997a). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hochreiter, S., & Schmidhuber, J. (1997b). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, *32*(10), 2513–2522. https://doi.org/10.1016/j.cor.2004.03.016

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

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. *2017 International Joint Conference on Neural Networks (IJCNN)*, 1419–1426. https://doi.org/10.1109/IJCNN.2017.7966019

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, *42*(4), 2162–2172. https://doi.org/10.1016/j.eswa.2014.10.031

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., Blunt, H., Brigham, T., Chang, S., Clark, J., Conway, A., Couban, R., de Kock, S., Farrah, K., Fehrmann, P., Foster, M., Fowler, S. A., Glanville, J., … Young, S. (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, *10*(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Small, H. (1973). Co□citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. https://doi.org/10.1002/asi.4630240406

Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, *12*(2), 430–435. https://doi.org/10.1016/j.joi.2018.03.006

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer New York. https://doi.org/10.1007/978-1-4757-3264-1

Wilson, G. T. (2016). Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978□1□118□67502□1. *Journal of Time Series Analysis*, *37*(5), 709–711. https://doi.org/10.1111/jtsa.12194

Xie, Y., Bian, C., Murphey, Y. L., & Kochhar, D. S. (2017). An SVM parameter learning algorithm scalable on large data size for driver fatigue detection. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8. https://doi.org/10.1109/SSCI.2017.8285380

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0