

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

Summer 3-22-2021

Bibliometric Survey for Stock Market Prediction using Sentimental Analysis and LSTM

Pooja Bagane
poojabagane@gmail.com

Nimit Mehta Mr
Symbiosis International University, nimumehta83@gmail.com


Parth Kakde Mr
Symbiosis Law School, parthkakade@gmail.com

Nisarg Bramhbhatt Mr
Symbiosis International University, brahmbhatt131@gmail.com

Ishansh Sahni Mr
Symbiosis International University, sahniishansh@gmail.com

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>

 Part of the [Computer and Systems Architecture Commons](#), and the [Library and Information Science Commons](#)

Bagane, Pooja; Mehta, Nimit Mr; Kakde, Parth Mr; Bramhbhatt, Nisarg Mr; Sahni, Ishansh Mr; and Kotrappa, Sirbi Dr, "Bibliometric Survey for Stock Market Prediction using Sentimental Analysis and LSTM" (2021). *Library Philosophy and Practice (e-journal)*. 5335.
<https://digitalcommons.unl.edu/libphilprac/5335>

Authors

Pooja Bagane, Nimit Mehta Mr, Parth Kakde Mr, Nisarg Bramhbhatt Mr, Ishansh Sahni Mr, and Sirbi Kotrappa Dr

Bibliometric Survey for Stock Market Prediction using Sentimental Analysis and LSTM

Pooja Anil Bagane¹, Nimit Mehta², Parth Kakade³, Nisarg Brahmhatt⁴, Ishansh Sahni⁵, Dr. S. Kotrappa⁶

¹Research Scholar, Visvesaraya Technological University, Belgaum, Karnataka- India. Assistant Professor at Department of Computer Science, Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.

Email: poojabagane@gmail.com

^{2,3,4,5}Student, Computer Science, Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.

⁶Professor, Department of Computer Science & Engineering, KLE Dr. MSSCET, Belgaum, Karnataka- India.

ABSTRACT

Creating an overview of the flow fundamentals of the global market trading and sublimation of equities into a superimposed system of economic agendas. Furthermore this leads to dynamic overlapping with the current technological advancements to create a platform for information exchange and inculcations. This created a new field of access points where we could enhance and analyse the data available and create an interface to predict the rise and fall trends involved with the stock market. These help create a sense of control and format over the public personification over the economic impacts and use social media and involve discrete approaches towards a smoother economic establishment. Here we review the work in the field of the aforementioned workspace and evaluate the established agendas.

1. INTRODUCTION

Asymptotically looking into the subject analysis we find ourselves into a dynamic that has had a splendid outlook over the recent times with greater amounts of funding and basis being provided onto the conceptual advancement to be seen. We make a greater stride in evaluating the various footnotes over the advancement studies set forth and aim to create a greater granular approach on the impactful study topics and correspondence growth.

There exist various fundamental stipends to measure and justify the process of involvement of the community towards various avenues in the study platform and we notice and create a helpful architectural platform of approachment where we can understand and learn the field of study and enhance the existential efforts.

First we generate an access point where we can take a deep dive into the functional matter and published matter made available to the global community. Furthermore we generate a specific and fixed class of data with a limited class size and create a pinpoint vibrancy and attachment towards the encapsulation.

Later we enhance the information onto a readable level and create an interactive diagrammatic approach of deep dive exclusions. Furthermore this creates a legibility to the data offered and opens an easement in understanding and interpreting data to a wider range of readers and creates a greater aesthetic approach.

2. PRELIMINARY DATA COLLECTION

In this bibliometric view of sentimental analysis traits in the stock market, we further noticed that we required a database network that could provide a more significant deal of insight over the existing matter available. We set forth by creating an access space of the SCOPUS tool in our arsenal at the library database access point to create a strong foundation. We sought out this model in the timeframe of March 2021.

2.1 Momentous Keywords

The fundamental keywords used in the study's underlying deep dive were: "Stock Market Analysis" and "Machine Learning." The secondary keywords pertaining throughout were "Sentimental Analysis," "LSTM," and "Social Media Analytics."

Table 1: Planned search tactic for Keywords

Primary Keywords (AND)	"Stock Market Prediction" and "LSTM"
Secondary Keywords (OR)	"Sentimental Analysis"

2.2 Initial Search Results

One hundred fifty-nine publications generated for the Initial Search Query. These publications are in various languages like English, Chinese and Turkish. Table 2 shows trends in publishing languages—the majority number of publications in English Language only.

Table 2: Trends in publishing language

Publication Type	Publication Count
English	156
Chinese	2
Turkish	1
Total	159

Source: <http://www.scopus.com> (March 2021)

A study has been published in journal papers, articles, book chapters, conference proceedings, etc., for this survey. The researchers in Stock Market Prediction using Sentimental Analysis and LSTM have publicized recent papers in conferences. 56% of conference papers and 37.1 % of journal articles were there. (Figure 1).

Documents by type

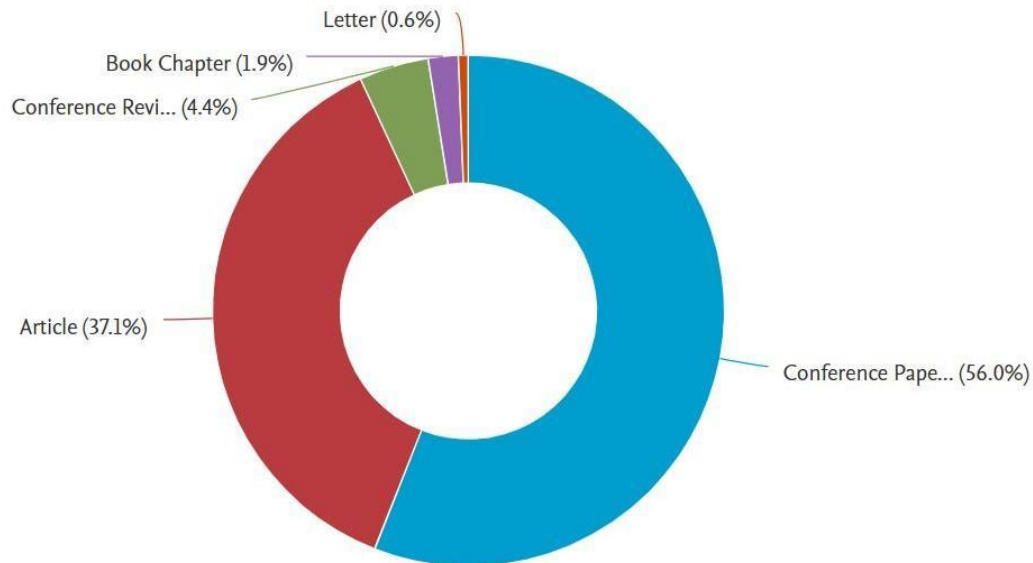


Figure 1: Analysis by Publication Type

Source: <http://www.scopus.com> (March 2021)

2.2 Highlights of Elementary data

An initial investigation is done based on the keywords that extracted 159 different types of publications from 2004 to 2020 in Stock Market Prediction using Sentimental Analysis and LSTM. The number of publications per year is shown in table 3. Analysis based on the number of publications per year is shown in figure 2. The majority of work was published in 2020, while no paper was published in 2014.

Table 3: Publication count per year

Year	Publication Count	Year	Publication Count
2021	15	2016	2
2020	61	2015	1
2019	49	2014	0
2018	17	2013	1
2017	12	2012	1

Source: <http://www.scopus.com> (March 2021)

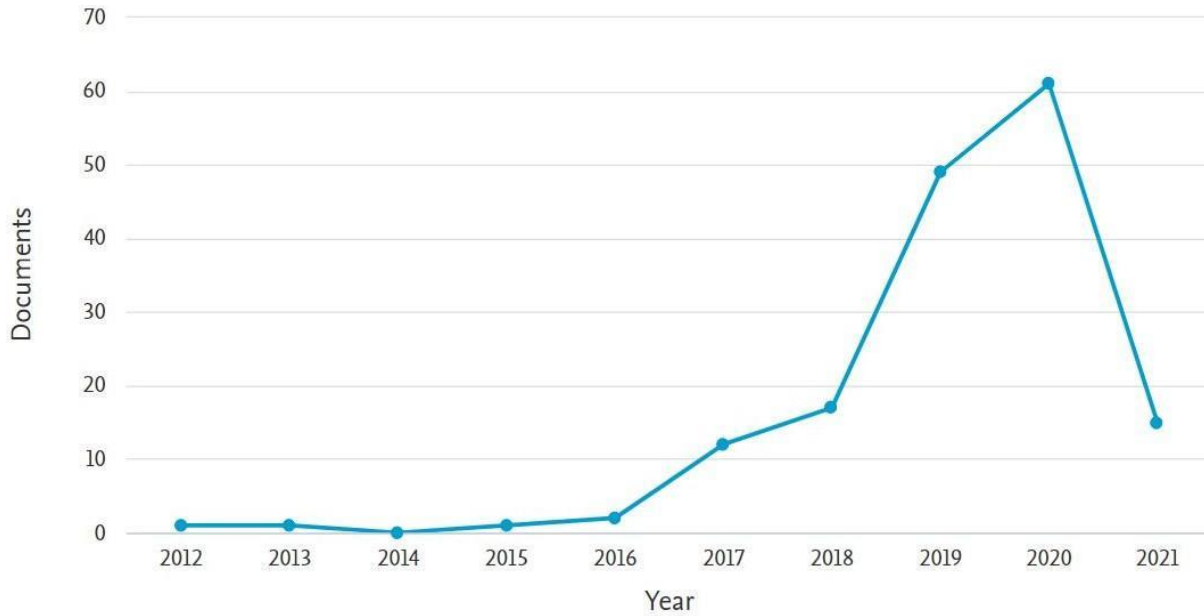


Figure 2: Analysis of Publication by Year

2.3 Geographical Region Analysis

The Geographical Region Analysis of 159 research papers found is shown in Figure 3. India has published the most number of documents with 49, while China is second with 40. Iran has published the least number of papers with a count of 4 among all considered countries. Figures 3 and 4 show different countries' contributions in published articles.

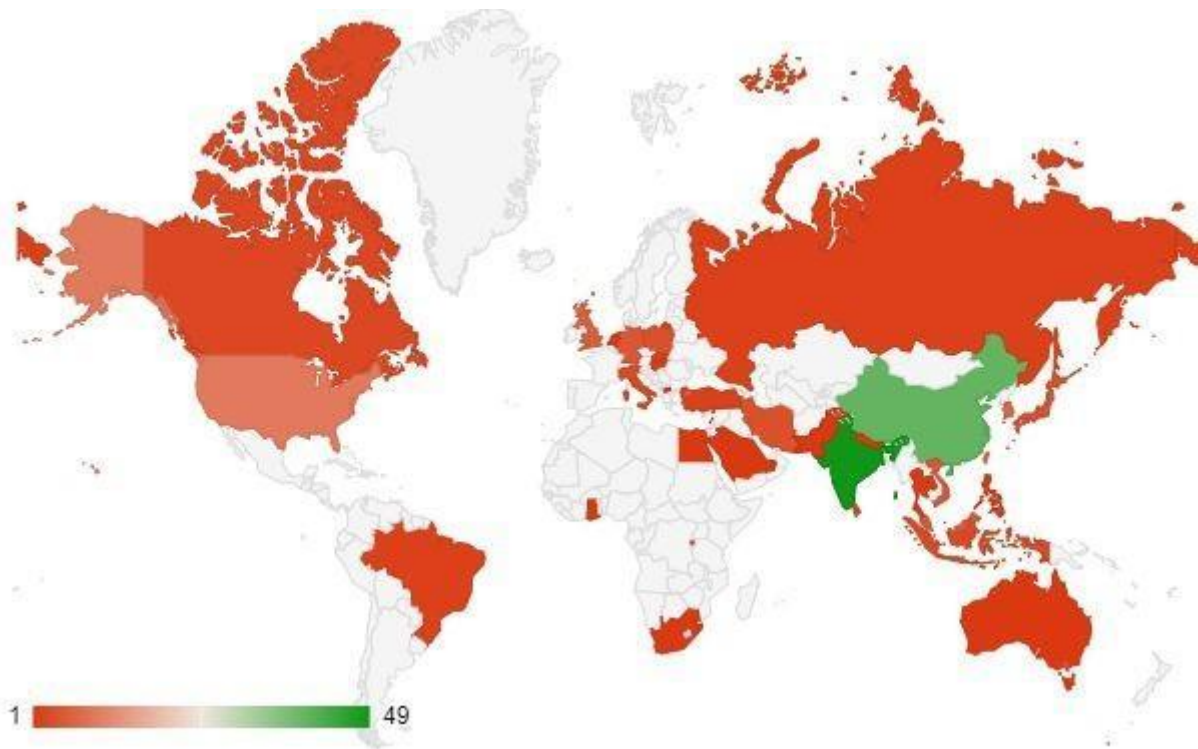


Figure 3: Research on Stock Market Prediction at different Geographic Locations

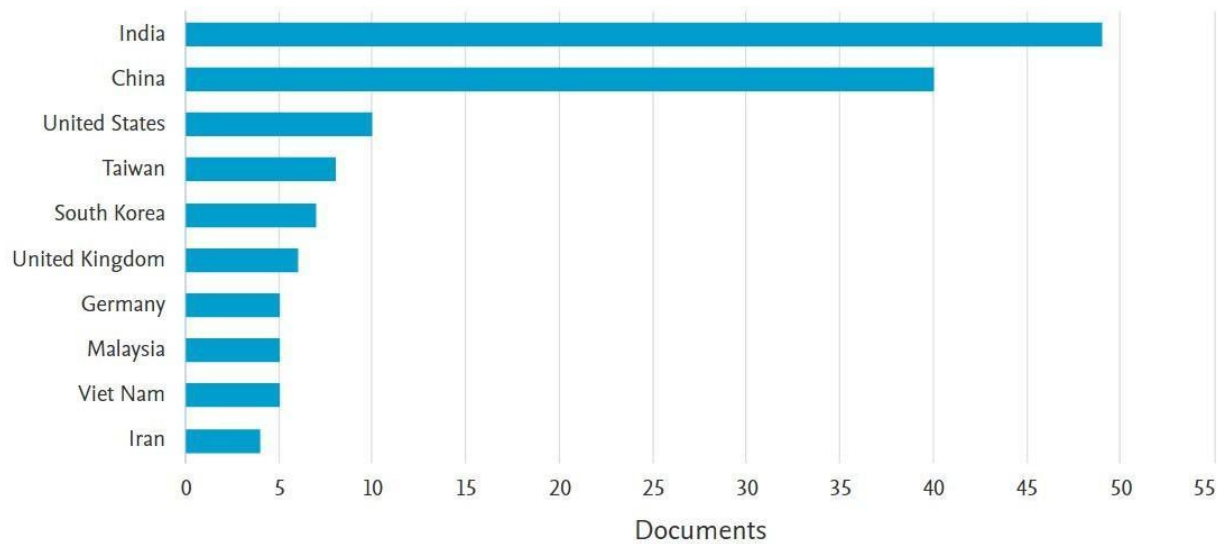


Figure 4: Analysis by country / Region

Source: <http://www.scopus.com> (March 2021)

2.4 Analysis based on Subject area

Figure 5 shows the classification of research papers based on the subject area. From this figure, it can be seen that most of the research work is done in the area of Computer Science, Engineering followed by Mathematics. The least amount of research is done in economics, Social Science followed by Material Science.

Documents by subject area

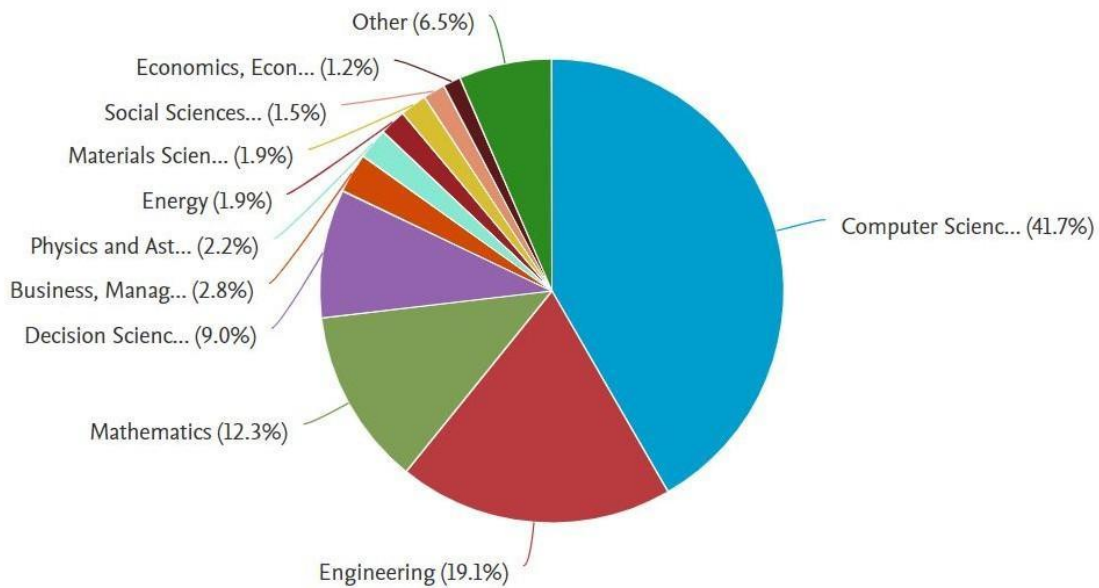


Figure 5: Analysis of Publications by Subject Area

Source: <http://www.scopus.com> (March 2021)

2.5 Analysis based on Affiliation

Figure 6 shows the contribution of various universities in publishing research work in Stock Market Prediction using Sentimental Analysis and LSTM. Chinese Universities have made significant contributions, as shown in the figure.

Documents by affiliation

Compare the document counts for up to 15 affiliations.

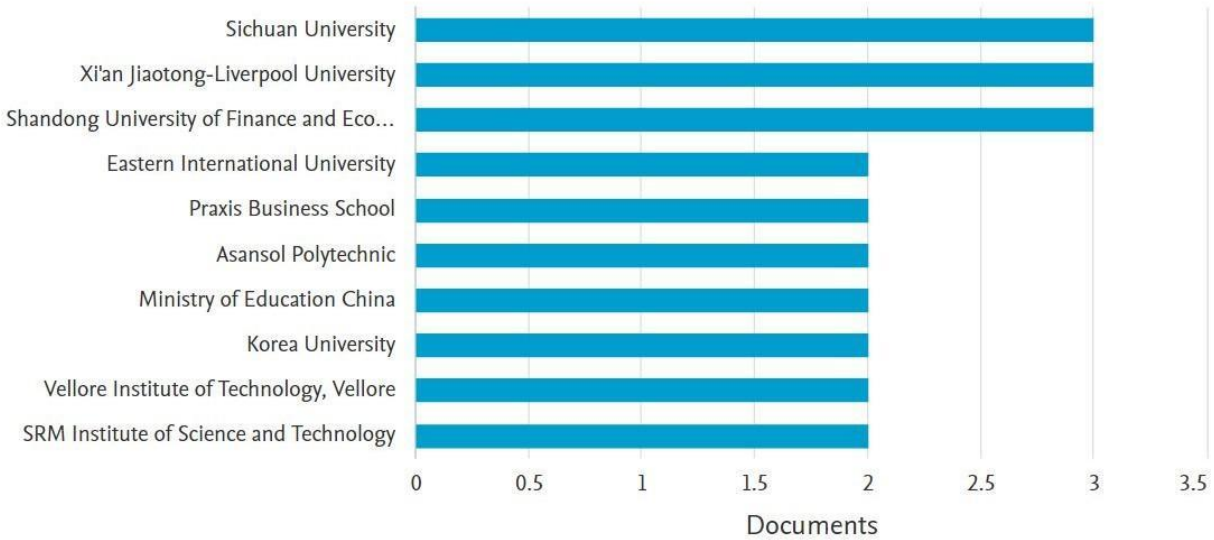


Figure 6: Analysis based on affiliations for publications

Source: <http://www.scopus.com> (March 2021)

2.6 Analysis based on Sources

Figure 7 shows the analysis of documents by sources. Most of the research work was published in Lecture Notes in Advances in Intelligent Systems and Computing, and Lecture Notes in Computer Science, including Subseries Lecture Notes in Artificial Intelligence Lecture, Notes in Bioinformatics. Communication in Computer and Information Science has minor publications in Stock Market Prediction followed by Expert Systems with Applications.

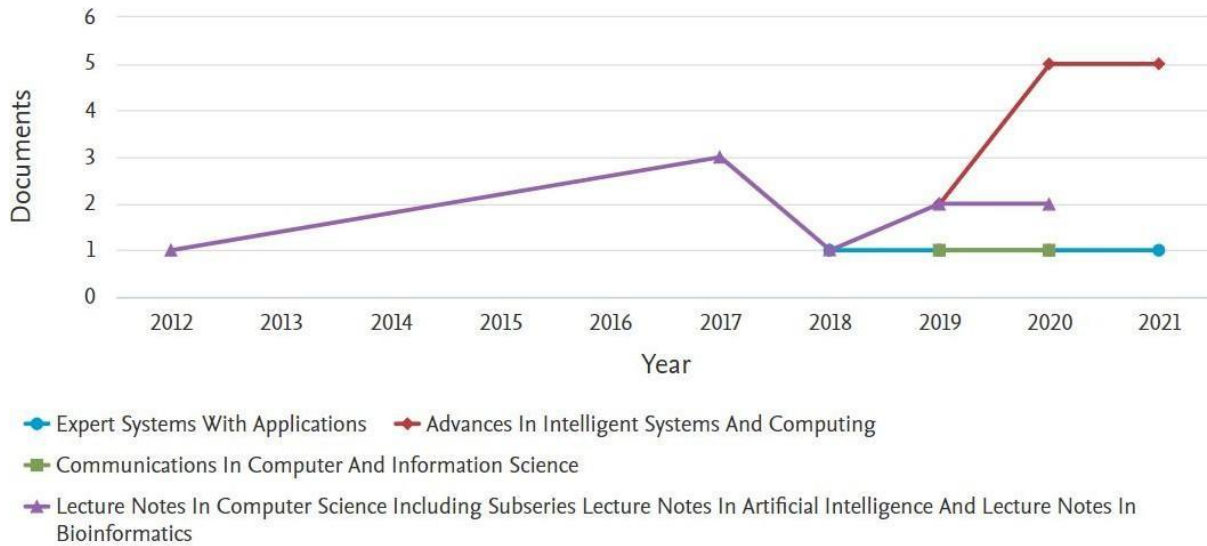


Figure 7: Analysis of documents by sources

Source: <http://www.scopus.com> (March 2021)

2.7 Analysis based on Funding Sponsor

Figure 8 shows the analysis of documents by funding sponsors. National Natural Science Foundation of China has given major funded sponsorship for the research in Stock Market Prediction using Sentimental Analysis and LSTM. In contrast, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Central South University, Center of Innovation Program, Celal Bayar Univeritesi, and Binus University has the least contribution.

Documents by funding sponsor

Compare the document counts for up to 15 funding sponsors.

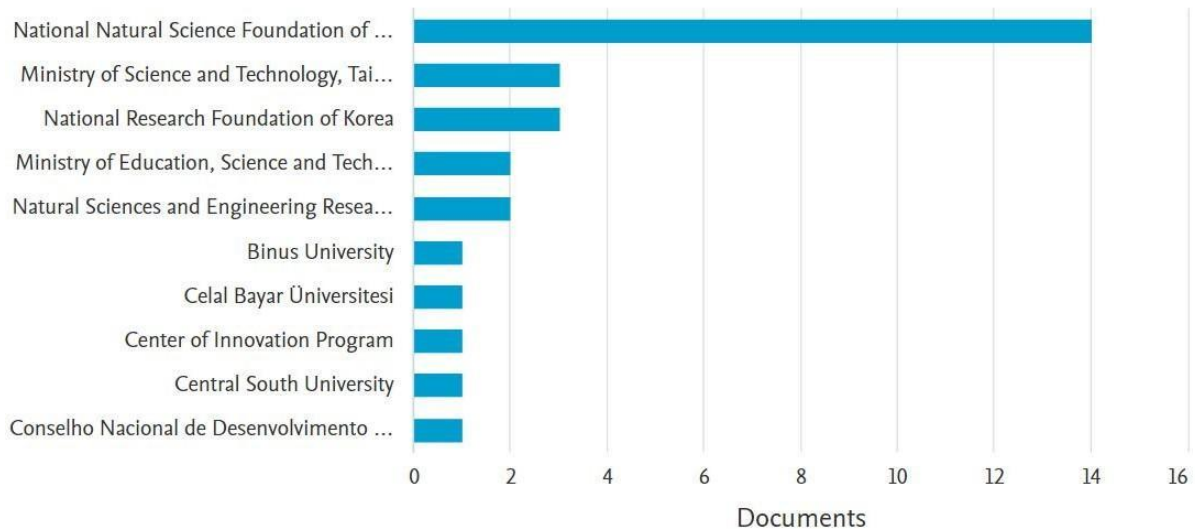


Figure 8: Analysis of documents by funding Sponsor

Source: <http://www.scopus.com> (March 2021)

2.8 Analysis based on Publications per author

Key authors contributing to Stock Market Prediction using Sentimental Analysis and LSTM are depicted in figure 9. The first ten authors were considered from the available accessed data from the Scopus database.

Documents by author

Compare the document counts for up to 15 authors.

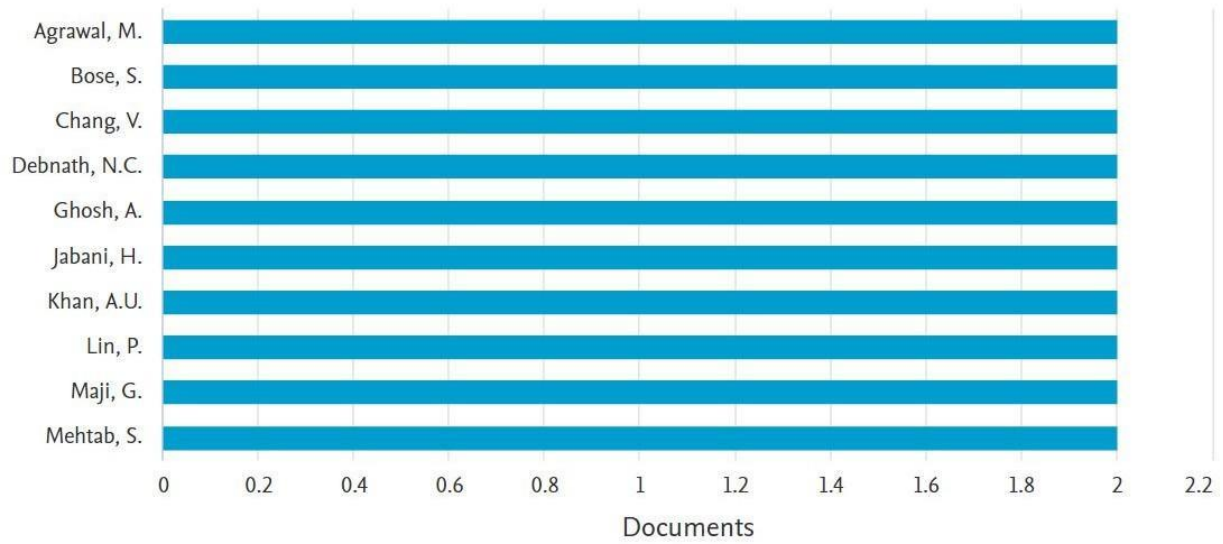


Figure 9: Analysis based on number of publications per author

Source: <http://www.scopus.com> (March 2021)

2.9 Citation Analysis

Table 4 shows overall citations for publication per year in cryptanalysis of block ciphers used in cybersecurity—one thousand two hundred fifteen sources for 159 publications to date. Titles of the top papers and citations received to them until the date of the data extracted for this research is in Table 5.

Table 4: Analysis of citations per year in the area of Stock Market Prediction

<2017	2017	2018	2019	2020	2021	subtotal
6	7	73	272	697	166	1215

Table 5: A citation analysis of top publications in Stock Market Prediction

Document Title	<2017	2017	2018	2019	2020	2021
Deep learning with extended short-term memory networks for financial market predictions	6	7	73	272	697	166
Stock market's price movement prediction with LSTM neural networks	0	0	9	77	176	41
Financial time series forecasting model based on CEEMDAN and LSTM	0	0	18	50	73	12
NSE Stock Market Prediction Using Deep-Learning Models	0	0	0	10	55	10
Artificial neural networks architectures for stock price prediction: Comparisons and applications	0	0	1	18	39	15
ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module	0	3	10	13	22	8
Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets	0	0	0	11	31	8

Evaluation of Bidirectional LSTM for Short and Long-Term Stock Market Prediction	0	0	3	11	21	2
LSTM neural network with emotional analysis for prediction of stock price	0	0	3	6	22	4
Exploring mutual information-based sentiment analysis with kernel-based extreme learning machine for stock prediction	0	0	4	6	18	2
An innovative neural network approach for stock market prediction	0	0	8	10	6	2
An ensemble of LSTM neural networks for high-frequency stock market classification	0	0	0	5	15	5
Novel volatility forecasting using deep learning–Long Short Term Memory Recurrent Neural Networks	0	0	0	0	17	4
DeepLOB: Deep convolutional neural networks for limit order books	0	0	0	0	17	1
Advanced Markov-based machine learning framework for making adaptive trading system	0	0	0	3	14	1
Forecasting stock prices with long-short term memory neural network based on attention mechanism	0	0	0	6	11	0
Robust and adaptive online time series prediction with long short-term memory	0	0	0	0	10	5
Model and forecast stock market behavior integrating investor sentiment analysis and transaction data.	0	0	1	3	9	1
Improving stock closing price prediction using recurrent neural network and technical indicators	0	0	5	5	3	0
Study of stock prediction based on social network	0	0	0	1	7	3
Stock closing price prediction based on sentiment analysis and LSTM	5	1	3	2	0	0

3. NETWORK ANALYSIS

The correlation between the various entities can be represented by using network analysis. This analysis was carried out by using different tools like Gephi, Sciencescape. Both are free software available for everyone and used for clustering and manipulation of the available data. By combining other parameters from extracted data of Scopus are used for creating the clusters shown below figures. Fruchterman Reingold's layout was used along with manual adjustments for the design while clustering.

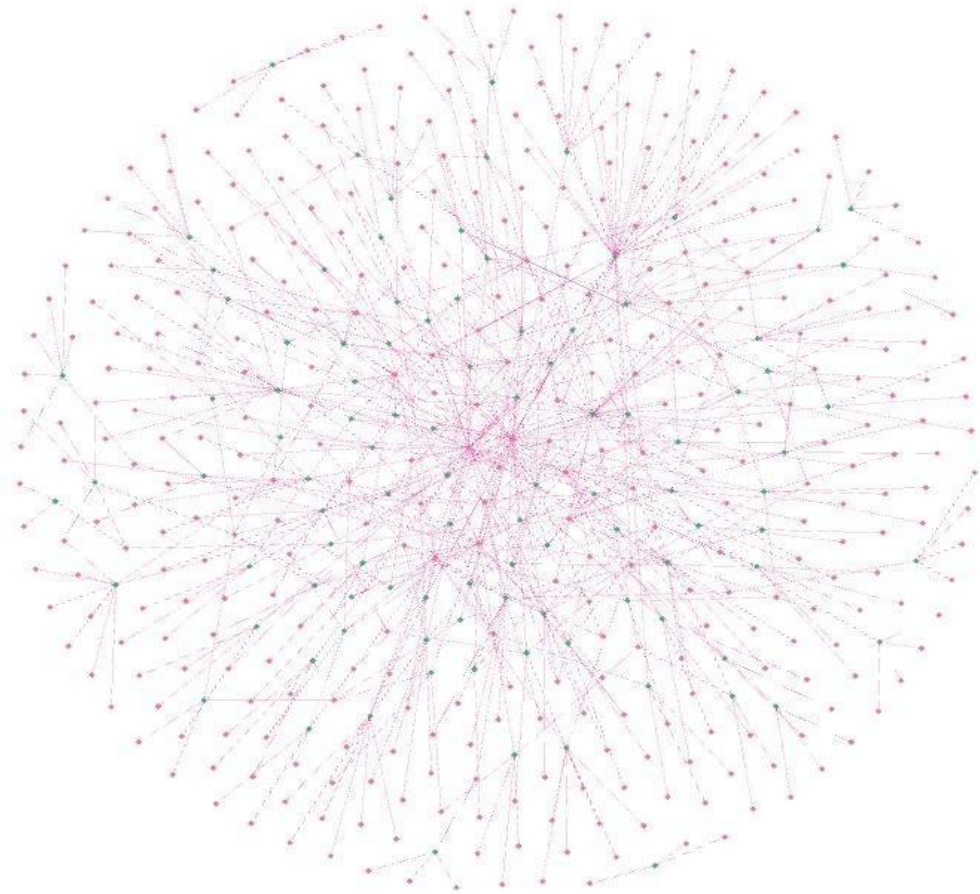


Figure 10: Cluster of Author Keyword and Source Titles

Figure 10 shows the analysis for Cluster of Author Keyword and Source Titles. 783 Nodes and 910 Edges are used to represent authors keywords and source title. Some nodes are removed by the science scope tool as they are disconnected from the network.

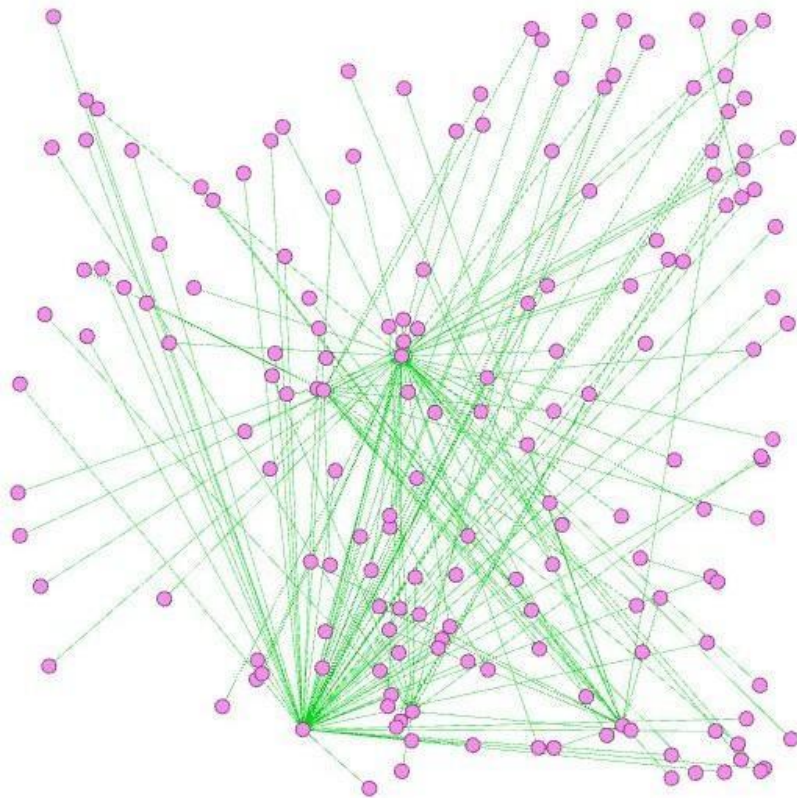


Figure 11: Analysis of Publication Title and Publication year

A cluster of Publication Title and Publication Year is shown in Figure 11. 240 Nodes and 220 Edges are used to represent Publication Title and Publication Year.

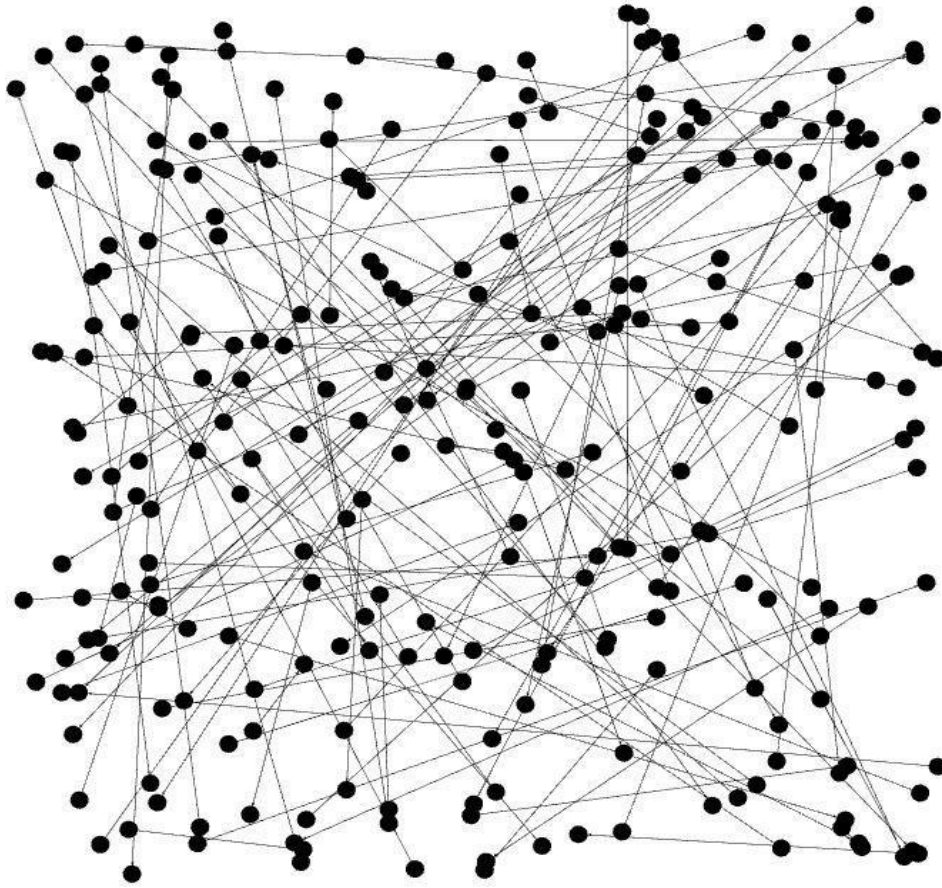


Figure 12: Cluster of Authors and Keywords

A cluster of Authors and Keywords is shown in Figure 12. 368 Nodes and 188 Edges are used to represent authors and keywords.

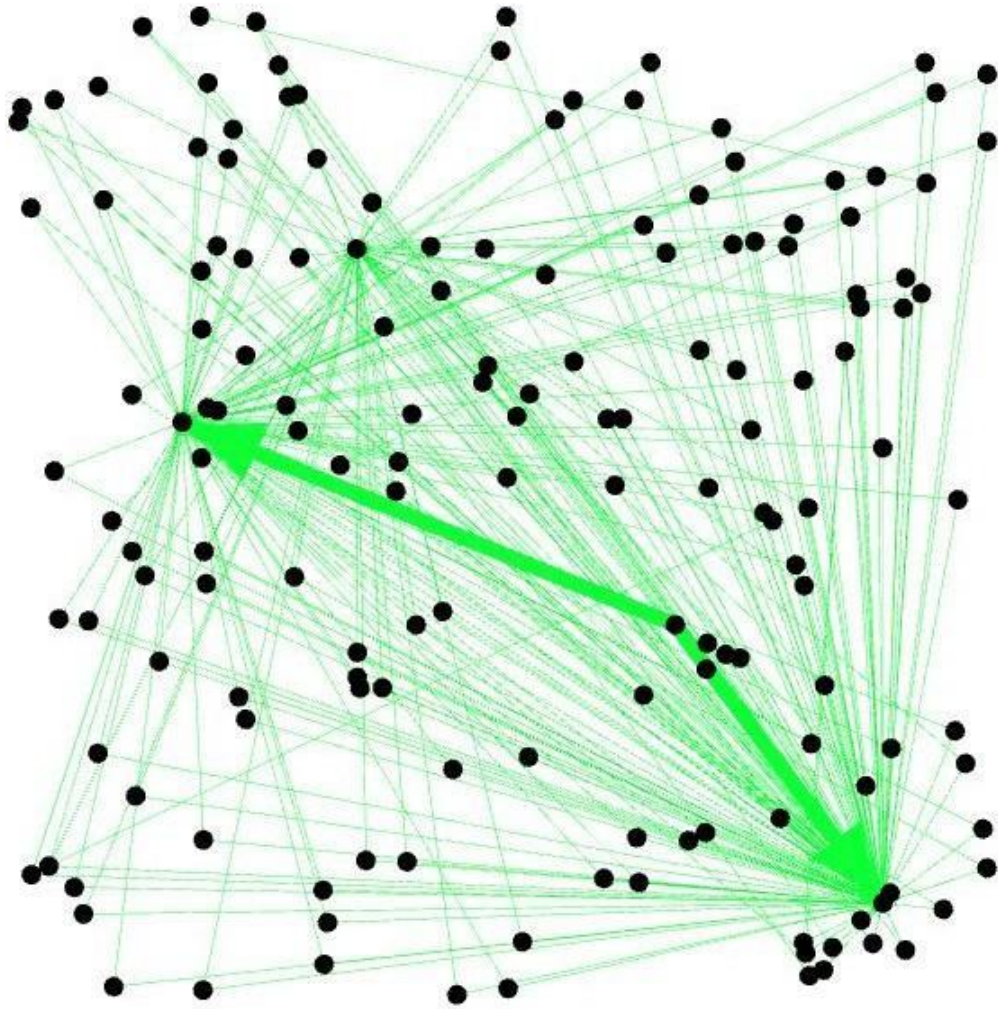


Figure 13: Cluster by Affiliation, Language, and Source Type

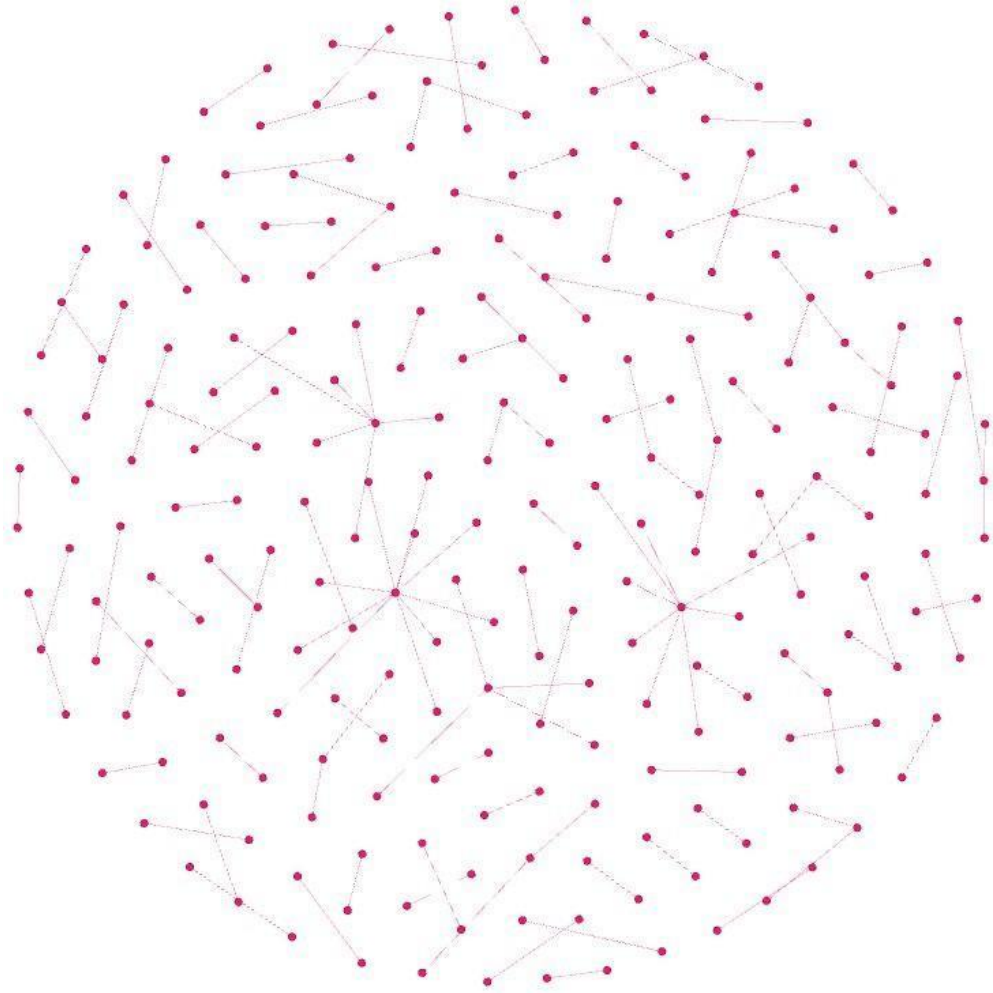


Figure 14: Cluster by Source Title and Author's Keywords

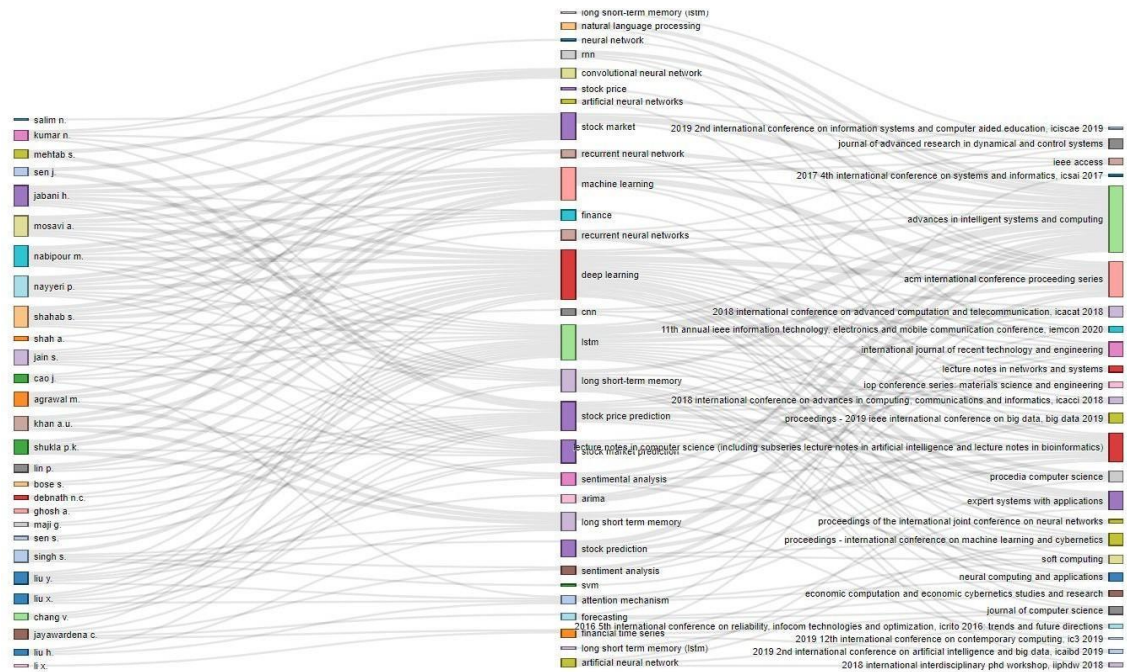


Figure 15: Relation of Authors, Keywords and Journals by Sankey Diagram

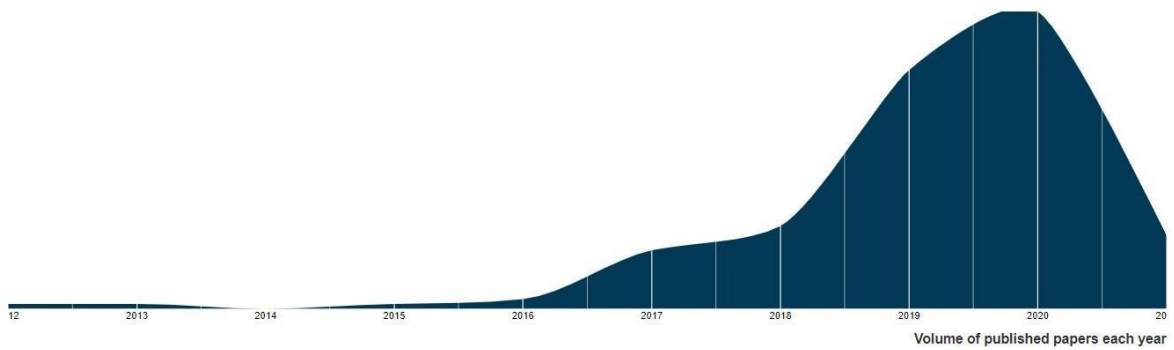


Figure 16: Analysis of Publications over time

4. LIMITATIONS OF PRESENT STUDY

This bibliometric study considers just Scopus-based publications resulting in the blending of keywords used by researchers. There are many databases like PubMed, Web of Science, and Google Scholar for research publication which are not analyzed during this bibliometric survey. Different databases give different statistics of citations for publication. In this study, sources for documents are also counted for the Scopus database only.

5. CONCLUSIONS

By doing the bibliometric study, it can be found that a very high number of researchers are working in the field of Stock Market Prediction using Sentimental Analysis and LSTM. It can also be seen from the study that the interest of researchers in this subject has increased over the last few years. As the research work will increase, it is expected that the work's accuracy will also increase.

6. REFERENCES

1. Nelson, D.M.Q., Pereira, A.C.M., De Oliveira, R.A. Stock market's price movement prediction with LSTM neural networks (2017) Proceedings of the International Joint Conference on Neural Networks, 2017-May, art. no. 7966019, pp. 1419-1426.
2. Wang, F., Zhang, Y., Rao, Q., Li, K., Zhang, H. Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction (2017) Soft Computing, 21 (12), pp. 3193-3205.
3. Zhuge, Q., Xu, L., Zhang, G. LSTM neural network with emotional analysis for prediction of stock price (2017) Engineering Letters, 25 (2), pp. 167-175.
4. Althelaya, K.A., El-Alfy, El.-S.M., Mohammed, S. Evaluation of Bidirectional LSTM for Short and Long-Term Stock Market Prediction (2018) 9th International Conference on Information and Communication Systems, ICICS 2018, 2018-January, pp. 151-156.
5. Zhou, X., Pan, Z., Hu, G., Tang, S., Zhao, C. Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets (2018) Mathematical Problems in Engineering, 2018, art. no. 4907423.
6. Baek, Y., Kim, H.Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module (2018) Expert Systems with Applications, 113, pp. 457-480.
7. Di Persio, L., Honchar, O. Artificial neural networks architectures for stock price prediction: Comparisons and applications (2016) International Journal of Circuits, Systems and Signal Processing, 10, pp. 403-413.

8. Cao, J., Li, Z., Li, J. Financial time series forecasting model based on CEEMDAN and LSTM (2019) *Physica A: Statistical Mechanics and its Applications*, 519, pp. 127-139.
9. Hiransha, M., Gopalakrishnan, E.A., Menon, V.K., Soman, K.P. NSE Stock Market Prediction Using Deep-Learning Models (2018) *Procedia Computer Science*, 132, pp. 1351-1362.
10. Fischer, T., Krauss, C. Deep learning with long short-term memory networks for financial market predictions (2018) *European Journal of Operational Research*, 270 (2), pp. 654-669.
11. Liu, Y. Novel volatility forecasting using deep learning–Long Short-Term Memory Recurrent Neural Networks (2019) *Expert Systems with Applications*, 132, pp. 99-109.
12. Borovkova, S., Tsiamas, I. An ensemble of LSTM neural networks for high-frequency stock market classification (2019) *Journal of Forecasting*, 38 (6), pp. 600-619.
13. Pang, X., Zhou, Y., Wang, P., Lin, W., Chang, V. An innovative neural network approach for stock market prediction (2020) *Journal of Supercomputing*, 76 (3), pp. 2098-2118.
14. Jin, Z., Yang, Y., Liu, Y. Stock closing price prediction based on sentiment analysis and LSTM (2020) *Neural Computing and Applications*, 32 (13), pp. 9713-9729.
15. Chen, Z., Du, X. Study of stock prediction based on social network (2013) *Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013*, art. no. 6693438, pp. 913-916.
16. Gao, T., Chai, Y. Improving stock closing price prediction using recurrent neural network and technical indicators (2018) *Neural Computation*, 30 (10), pp. 2833-2854.
17. Zhang, G., Xu, L., Xue, Y. Model and forecast stock market behavior integrating investor sentiment analysis and transaction data (2017) *Cluster Computing*, 20 (1), pp. 789-803.
18. Yang, H., Pan, Z., Tao, Q. Robust and adaptive online time series prediction with long short-term memory (2017) *Computational Intelligence and Neuroscience*, 2017, art. no. 9478952.
19. Qiu, J., Wang, B., Zhou, C. Forecasting stock prices with long-short term memory neural network based on attention mechanism (2020) *PLoS ONE*, 15 (1), art. no. e0227222, .
20. Rundo, F., Trenta, F., Di Stallo, A.L., Battiato, S. Advanced Markov-based machine learning framework for making adaptive trading system (2019) *Computation*, 7 (1), art. no. 4.
21. Bagane, Pooja and Kotrappa, Sirbi, "Bibliometric Survey for Cryptanalysis of Block Ciphers towards Cyber Security", *Library Philosophy and Practices*, (2020).