

A Bibliometric Review of Stock Market Prediction: Perspective of Emerging Markets

Arjun Remadevi Somanathan^{1*}, Suprabha Kudigrama Rama²
^{1,2}National Institute of Technology Karnataka, Karnataka 575 025, India

Abstract – The objective of the paper is to identify predictive models in stock market prediction focusing on a scenario of the emerging markets. An exploratory analysis and conceptual modelling based on the extant literature during 1933 to 2020 have been used in the study. The databases of Web of Science, Scopus, and JSTOR ensure the reliability of the literature. Bibliometrics and scientometric techniques have been applied to the retrieved articles to create a conceptual framework by mapping interlinks and limitations in past studies. Focus of research is hybrid models that integrate big data, social media, and real-time streaming data. Key finding is that actual phenomena affecting stock market sectors are diverse and, hence, limited in generalization. The future research must focus on models empirically validated within the emerging markets. Such an approach will offer an insight to analysts and researchers, policymakers or regulators.

Keywords – Bibliometrics, emerging markets, stock market prediction, systematic review.

I. INTRODUCTION

Research in the field of stock market goes back to the early 1930s. Apart from concrete theories, most studies have contributed to merging and sub-diverging of academic disciplines and varied findings. The earliest real stock market recorded in human history was Amsterdam stock exchange in 1602. Securities or shares/stocks that are comprised of different companies or firms can be grouped and tracked under a common notion called a stock index. Before the turn of the 21st century, this very definition has paved the way to different notions with the rise of electronic financial markets [1]. It is human psychology that investors and other stakeholders in this market wish to get profit regardless of actual economies involved in the transaction environment. From the perspective of academic disciplines of Statistics and Finance, researchers started looking into how the stock market behaves or prices can be predicted [2]. Three key theories have been developed from the earliest studies: random walk hypothesis (RWH), efficient market hypothesis (EMH), and noisy market hypothesis. In the current study, some major contributions are as follows.

1. The research growth of stock market prediction models documented from 1933 to 2020 are extracted from Web of Science and Scopus.

2. The authors analyse the major factors, such as most productive and influential researchers, highly cited journals.
3. Detailed synthesis of top papers on the basis of the total citations received till date from both databases.
4. The visualization of the highly influential works and inter-referencing in this field are displayed using citation data.
5. In addition to bibliometric methods, theoretical implications are discussed from the influential papers on the practice of emerging markets to provide the areas for future studies.

Bibliometrics or scientometric analysis is the research field to analyse current trends in the literature of a specific research field and delivers procedures and impetuses for future research. Earlier, [3] have provided precise explanation of a bibliometric study. Similarly, one study by [4] has given the motivation to work in this area. Rest of the paper is organised as follows. Section 2 provides a review of literature. Section 3 introduces the research methodology. The framework is proposed in Section 4. Finally, conclusions are drawn in Section 5.

II. LITERATURE REVIEW

In the current study, two research questions are addressed. Research Question 1: Which are major predictive models for stock market prediction in emerging markets? This leads to the investigative question IQ1: Why is there a dearth of research in stock market forecasting models focused on sector/industrial categories in emerging markets? Research Question 2: What are emerging research themes of stock index forecasting in this scenario? The investigative question is IQ2: How can different theoretical frameworks of prediction models be validated in the Indian context?

In effort to answer these questions, a systematic but not exhaustive review is first carried out. Only scientific manuscripts are considered in the literature review. Any tertiary sources or industry reports etc. are not considered. Then Indian context works are compared on the basis of methodological variations, findings and contributions towards an overall research theme, and field of interest. Beginning with the theoretical perspectives, a random walk hypothesis (RWH) proposes that changes in stock prices have the same distribution

* Corresponding author e-mail: arjrs123@gmail.com

and are independent of each other; therefore, past movement or trend of a stock price or market cannot be used to predict its future movement [5]. As a Nobel winning work, an efficient market hypothesis (EMH) is consistent with the random walk theory, which states that asset prices fully reflect all available information in the market. A direct implication is that it is impossible to “beat the market” on a risk-adjusted basis since market prices should only react to new information. There are three variants such as ‘weak’, ‘semi-strong’ and ‘strong’ denoting a varying degree of this effect. “Noise” as a financial

concept affecting information was theorised by [5] and further it was developed into a coherent framework in the market environment by [6]. It contrasted the efficient-market hypothesis claiming that prices of securities were not always the best estimate of the true underlying value of firm. The study found that prices could be influenced by speculators and momentum traders, or insiders and institutions that often participated in market for reasons that were distinct to normal trade conditions.

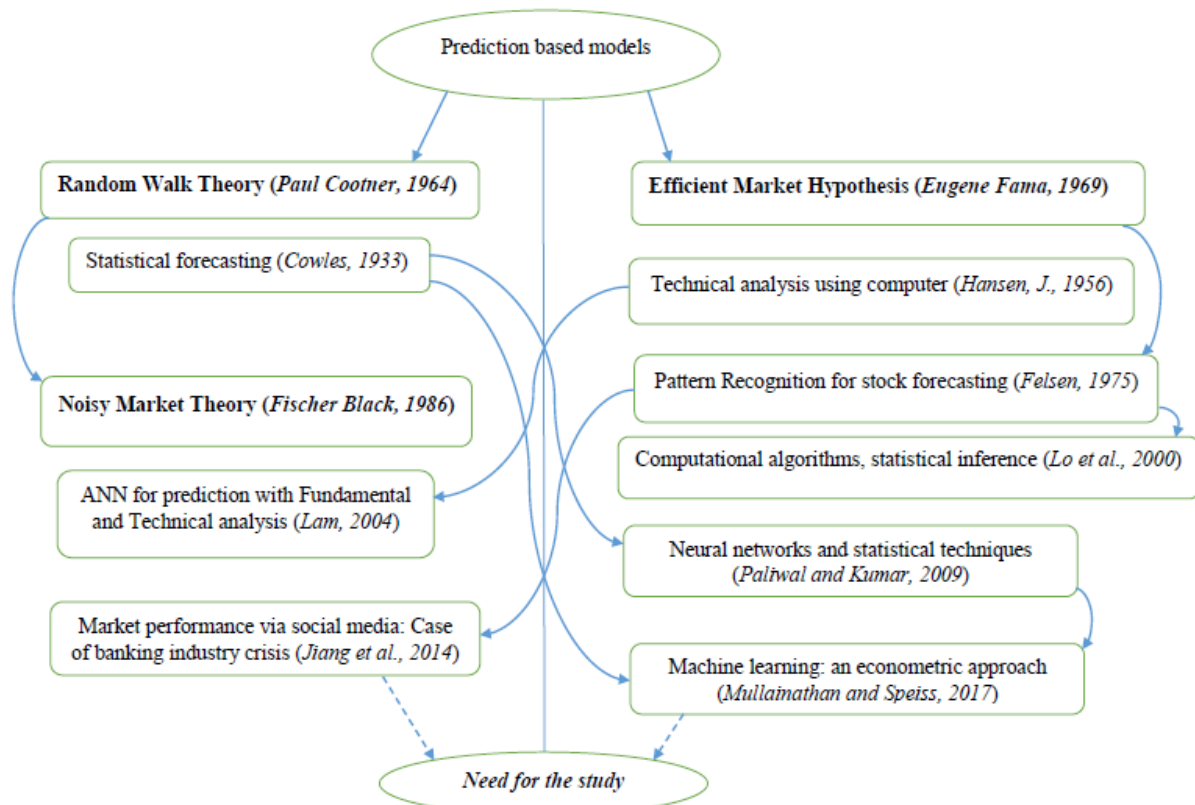


Fig. 1. Literature map (created by the authors).

Figure 1 depicts the mapping of the selected research literature from 1933 to 2019 along with theoretical mapping to studies. The interest in prediction models arose after 2000 and they were widely used in various areas such as business and accounting, finance, managerial applications, and operations research etc. In the case of stock market prediction, there is a lack of universal consensus as to which methods to adopt. Previously, studies shown recommending both parametric and non-parametric models for forecasting purposes.

Considering stock market as a dynamic system [7], there is a retrospection need of prediction as posed by [8], who expressed scepticism over relying on modelling in qualitative sense solely. The question of modelling with regard to quantification was proposed. Earlier, econometric theories were rigorous in mathematical formulation and widely used by analysts and stakeholders.

[2] quantified the financial services of stock investment prediction during 1929 to 1932 and statistically found that they

were worse than an average of common stocks annually. These results indicated that most of them were by chance rather than skilfully exhibited by analysts. With the development of digital computers and technologies, researchers started looking for applying them in practical problems. [9] mentioned the benefits of using computer in the stock market analysis from the perspective of technical analysis. The system was developed for the New York stock brokerage firm. Findings concluded that the categories of task should be done by an analyst; segregating of technical and fundamental indicator variables and complexities involved was assumed during the computation steps. The study suggested the type of analysis required on past history to determine validity, checking of other current theories of market analysis. In the later period, a technical analysis as a branch of active investment plan and research gained after [10] gave foundational ideas involved in the process.

During this period, investment management slowly began to evolve as an interdisciplinary field consisting of financial

analysts, engineers and industry professionals. For ex: [11] by using a decision model elaborated five theories intended for an investment analysis that could be applied to stock selection and market forecasting tasks. On the basis of experimental techniques using New York composite index datasets from 1970, it supported an earlier finding that an investment analysis required the processing of complex information patterns. Also, that the investment policy had to reflect a number of fundamental, psychological, technical, and other factors. Even at this stage, statistical methods retained their importance and were the primary evidence in suggestions.

A study [12] briefly reviewed the statistical classification methods of stock analysis and stated that efficient market theories tended to deny the possibility of a detailed analysis. Application of logit, probit, and tobit techniques for involving qualitative dependent variables was of importance.

Major issue of earlier studies was the elimination of vital fragment from population when choosing a priori groups for an analysis. From financial research literature, [13] gave an insight into the fundamental analysis and its roots that helps in investment decision making.

[14] in a seminal study solely surveyed a technical analysis, various smoothing techniques such as nonparametric kernel regression. They inferred an approach incorporating the essence of technical analysis to identify regularities in the time series of prices by extracting nonlinear patterns from noisy data. Daily returns of individual NYSE/AMEX and NASDAQ stocks from 1962 to 1996 were used along with bootstrapping and Monte Carlo simulation procedures. They concluded that a technical analysis could add value to the investment process.

[15] tested an experimented nonparametric model, such as ANN incorporating both technical and fundamental factors. The predictor attributes included 16 financial statement variables and 11 macroeconomic variables. Data of 364 S&P companies during the period of 1985–1999 were used. Results supported the ability of a neural network to significantly outperform the minimum benchmark based on a highly diversified investment strategy. Intuitively, financial statement variables and macroeconomic variables together cannot produce significantly higher returns than an average index and such integration better suits during economic recession.

[16] stated that neural network models were preferable when the relationship between variables was not known or was complex. Weakness of neural networks was a lack of interpretability of weights during the model building process. In this respect, a statistical model clearly stands out as it allows for the interpretation of coefficients of the variables. They opined that combining the features of both techniques enhanced overall prediction/ classification performance.

[17] conducted the innovative study that effectively deployed an experimental setup with methods to use social media data for stock modelling. They focused on a specific stock market, such as banking. Using methods such as an authorship analysis, topic analysis and sentiment analysis on social media, they developed a stock prediction regression model for the Bank of America. They found that a crisis affected social media discussion topics

and that different stakeholder groups/topics had distinct effects on market predictions.

[18] explained caveats and importance of prediction in estimation and policy modelling avenues. According to them, machine learning can also be used to build a benchmark for revisiting and testing well-known theories. The study also revealed that systematic variation captured in R^2 as a model fitness measure could not provide adequate answers to particular phenomena of interest.

III. RESEARCH METHODOLOGY

A. Strategy for Research

The Scopus database and Web of Science were used to extract the relevant manuscripts. Few articles were retrieved from the JSTOR in case of older publications according to the topic. The keyword used to retrieve results was “*Stock market index prediction*”. From the results obtained, the document type “Journal” was selected to filter such articles. Under this method, exploration of corpus of literature included 545 articles, grouped in 15 clusters using association strength of similarity and the total number of reference links was 16881. Figure 2 shows the bibliographic coupling visualization from Scopus data using VOSviewer. The blue colour dots indicate older research articles and yellow dots show more recent ones. The analysis technique follows the method by [19].

Finally, the most top 10 influential articles as per Google scholar metrics until Nov 2020 are:

- i. [20] Paper on Support vector model. 1628 citations.
- ii. [21] Study on effects of liberalization and emerging markets. 1640 citations (Interdisciplinary).
- iii. [22] Empirical study on Istanbul market prediction. 637 citations.
- iv. [23] Compared neural networks. 567 citations.
- v. [24] Brought difference in classification and estimation models. 454 citations.
- vi. [25] Designed optimization for back propagation improving generalizations. 426 citations.
- vii. [26] Tested adaptive fuzzy system for Istanbul market. 391 citations.
- viii. [27] Case study using a genetic algorithm for decision support. 367 citations.
- ix. [28] Fuzzy rule system. 355 citations.
- x. [29] Used a hybrid model. 340 citations.

As shown in Fig. 3, the journal Expert Systems with Applications had most articles that also had a high impact factor. It was followed by Physica A Statistical mechanics and its Applications journals not having articles prior of 2000. The overall trend shows a spike in 2009 due to the effect of global recession and renewed scholarly interest continuing to 2015.

There is a clear opportunity for Indian/emerging markets to test with innovative ideas, stimulate thought and practice to advance the field. Figure 4 shows density visualization.

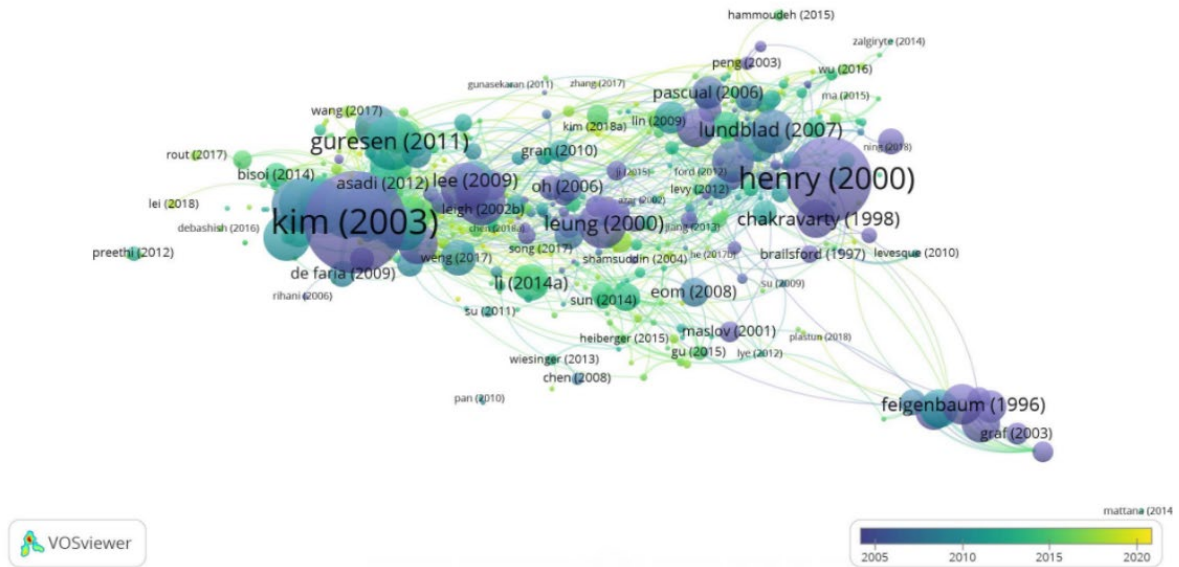


Fig. 2. Bibliographic coupling (source: VOSViewer).

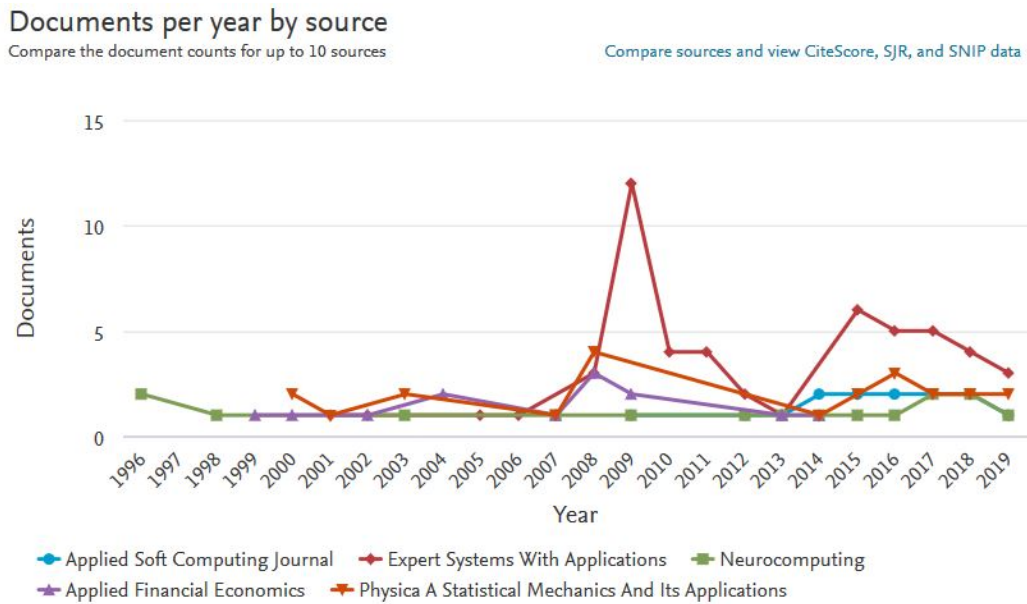


Fig. 3. Published research timeline (source: Scopus).

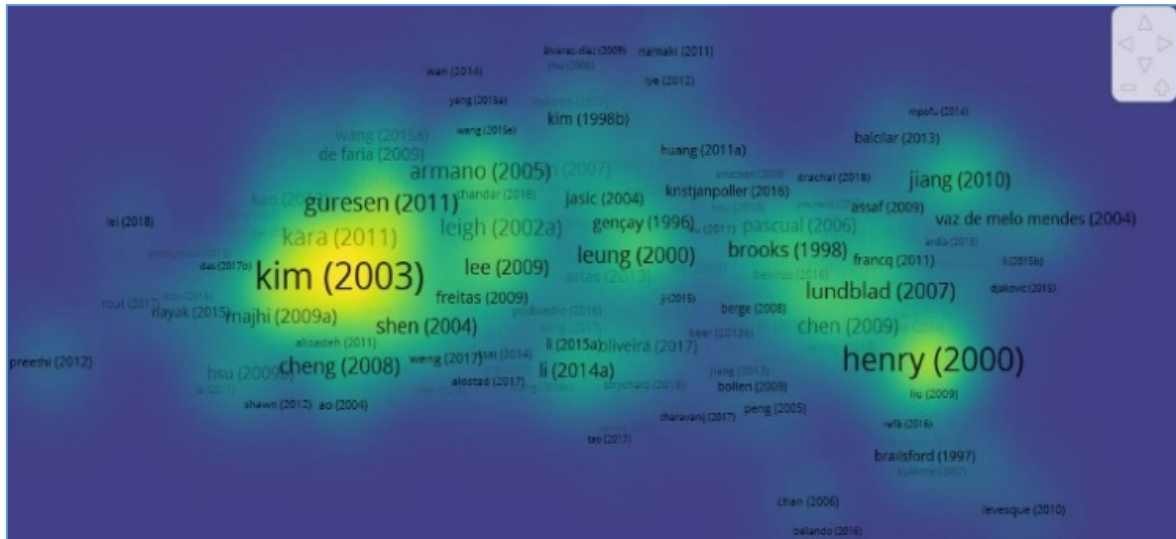


Fig. 4. Density visualization (source: Scopus).

Density visualization uses association strength, i.e. the similarity s_{ij} between two items i and j is calculated as follows:

$$s_{ij} = \frac{C_{ij}}{W_i W_j} \tag{1}$$

Here in Eq. (1), c_{ij} denotes the number of co-occurrences of items i and j and w_i and w_j denote either the total number of occurrences of items i and j or the total number of co-occurrences of items [19]. Here, the yellow nodes are high impact papers, blue are citing works.

B. Systematic Review of Articles

Table I explains top 3 articles thoroughly. Interestingly, [30] a major survey work did not find a single Indian empirical study. It may be noted that many influential articles found in the Web of Science also appear in Scopus results. Hence, they were omitted to avoid repetition while conducting review process. Findings support efficacy of neural network models over statistical methods though constant model updating is needed.

TABLE I
HIGHLY CITED RESEARCH WORKS

Author/ study	Objective/s	Methodology adopted	Sampling/ Variables	Findings
[20]	To forecast the direction of daily change in the Korean composite stock price index (KOSPI).	Support vector machine with Gaussian radial basis function. Experiments with artificial neural networks (ANN) using back propagation for case-based reasoning (CBR).	Sample of 2928 trading day's data from January 1989 to December 1998 (10 years). 12 technical indicators.	SVM performs better than CBR at 5 % statistical significance and has more generalizability.
[22]	Predicting the direction of the movement in the daily Istanbul Stock Exchange (ISE) National 100 Index.	Two models using ANN and SVM (support vector machines).	Daily closing price from 2 January 1997 to 31 December 2007. Application of 10 technical indicators derived from data.	Average performance of ANN model (75.74 %) is significantly better than SVM model (71.52 %).
[23]	To compare performance of ANN models in forecasting time series used in market values. Autoregressive Conditional Heteroscedasticity (ARCH) model, Generalized ARCH (GARCH) model, Exponential GARCH (EGARCH) model and Dynamic Architecture for ANNs (DAN2) are used.	Multilayer perceptron (MLP), dynamic architecture for ANN, GARCH-MLP models, EGARCH-MLP Models. Model performance is measured using mean square error (MSE) & mean absolute deviation (MAD).	Daily stock exchange rates of NASDAQ from 7 October 2008 to 26 June 2009. First 146 days are used for training and cross-validation and last 36 used for testing.	Classical ANN model MLP outperforms DAN2 and GARCH-MLP with a little difference. Longer periods over 17-day forecasting are hard by a simple technical analysis (only using previous index values or variables derived). Needs fundamental analysis (index with global, national economic, market & ratio analysis etc.).



Fig. 5. Subject classification of papers (source: Web of Science).

It is seen here that with the results found: 347, Sum of the Times Cited: 8078 and Average Citations per Item: 23.28. Figure 5 shows a Tree Map of academic research on stock market prediction. Bigger boxes show a higher number of articles. The leader is computer science and artificial intelligence (AI) with 227 papers. 16 papers are in statistics and probability area (Web of Science classification).

There is an increased interdisciplinary research work found in both databases. Further analysis is performed with Bibliometrics method like [19]. Number of articles for Bibliometrics coupling: 10, number of links: 23 and total link strength: 96. Research conducted focused on neuro-fuzzy systems [31] also in the forecasting models [32].

C. Overview of Indian Perspective Works

In Table II, influential works are described. Since this approach used a broader term of scope, Web of Science and Scopus databases were again researched with the keyword “stock market index prediction”. This narrowed down the articles that focused on particular index/demography rather than the whole market. Filtering to “India” was used in the field country for the Indian context. The retrieved articles were 3 in Scopus and 1 in Web of Science. One published study was common in both.

TABLE II
INDIAN PERSPECTIVE STUDIES

Sl. no	Authors	Objectives	Methods / Research theme	Findings	Limitations & Theory directions
1	[33]	Predicting the direction of movement in stock price index for Indian markets.	Machine learning.	Performance of all prediction models improve when technical parameters are represented as trend deterministic data.	Macro-economic variables such as currency exchange rates, inflation, government policies, interest rates etc. yet to be used. Adaptive market hypothesis may be used in model performance metrics.
2	[34]	Five methods combined to predict day’s closing price. Typical Price (TP), Bollinger Bands, Relative Strength Index (RSI), CMI and Moving Average (MA).	Using BSRCTB (Combinatorial Technique) of methods.	Data mining methods show that designing a user interface for prediction model is feasible.	Can be correct 50 % of the time significantly, but the algorithm still gives error consecutively. The algorithm generated both an increase and a decrease in predictions, but the results were not consistent. Since it does not work across spectrum of stocks, the behavioural investment theory can be tested in future works.
3	[35]	Predict the volatility of Bombay stock exchange.	Artificial neural networks.	Error increases in the validation period of model. Chances of over-parameterization.	Retraining of the prediction model suggested in every 6 months needs to incorporate structural changes within a specific market. Statistical learning theory.

D. Keyword Analysis

Following this method, Fig. 6 shows the visualization of keywords in research manuscripts. The keyword dataset was retrieved from the articles in Scopus (only Indian). Total of 147 articles were included in the analysis. The dataset had a total of 1222 keywords (comprised both author and index keywords). Minimum occurrences of keyword were fixed as 5. Therefore, 72 keywords met the threshold.

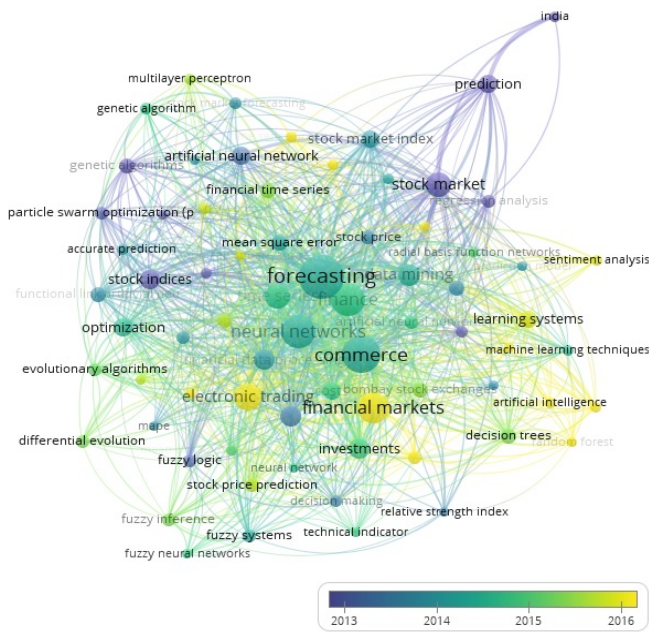


Fig. 6. Keyword analysis (source: VOSViewer).

Here, the yellow colour indicates the keyword in newer publications. Trend is in shift to financial markets, electronic trading, machine learning and artificial intelligence, multilayer perceptron, extreme learning machine, random forests, evolutionary algorithms etc. These models implement learning systems, sentiment analysis methods for predictions.

IV. THE PROPOSED FRAMEWORK

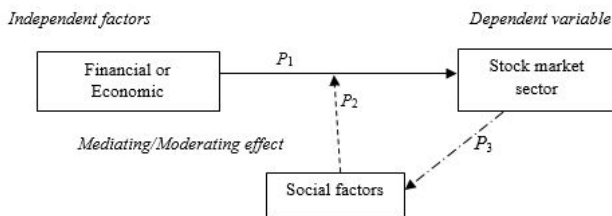


Fig. 7. Conceptual framework for future work (developed by the authors).

Figure 7 shows the proposed framework for evolving future research works. Here, instead of treating social factors as independent construct, studies must assume parameters that have a mediator effect while assuming a broader scale of stock prediction models of particular markets. Considering the lack of empirical studies in this area, such an approach provides new

implications for policy makers and industry practitioners as well. According to this framework, three hypothesized relations need to be subject to an empirical validation. Appendix I and II list the studies categorised based on factors surveyed with regard to research propositions.

Proposition 1 (P_1): The financial or economic factors have a significant and positive impact on the specific sector/industry of stock exchange of emerging market. Even though this relationship is heavily researched, there are lots of unanswered questions. A lot of intricacies and complexities for understanding such include regional differences, local economic factors propagated in media or news [36].

Proposition 2 (P_2): The social factors have a moderating or mediating effect on the relationship between financial conditions to a particular sector/industry in stock exchange of a country. This proposition holds much value in terms of bringing changes in the market landscape. It can be attributed due to the advent of technology, smartphone penetration or big data and social media influence [37]. It has been proven in earlier research with regard to social media platforms and their impact on stock market [38]–[40].

Proposition 3 (P_3): The stock market fluctuations in a specific industry have an observable impact on the social fabric. Very less research has been conducted in this dimension. Social factors may also be manifested in terms of technological progress [41] and corporate social responsibility (CSR) [42]. These propositions put forth a call for empirical work towards testing them as hypotheses in varied contexts. For example, the impact of disclosure by companies facilitates information aggregation and market reaction [43]. Method such as partial least squares (PLS) is also feasible [44]. Additionally, a data envelopment analysis (DEA) can aid in efficiency and performance estimation for investment decision making [45].

V. CONCLUSION

After the comparative analysis of foreign and Indian context, few inferences have been found critical. Subsequently, pitfalls of research in past Indian studies have been summarised. Social media, big data sources have not been used in most of framework studies, at least in major Indian works. Additionally, investor sentiment and group dynamics have been the primary factor of stock market outcomes that require more empirical Indian studies [46], [47]. The usage of hybrid models being rarely experimented gives mixed results, so they hold potential in the scenario of emerging markets [48]–[50] that consider the security aspects as well [51].

Using the aforementioned models, both quantitative and qualitative factors can be incorporated in model building and prediction purposes [52], [53]. Economic and policy implications of stock market models specific to sectors are rarely explored yet they can be developed. It is of interest and optimism that future researchers will combine the existing tools, big data sources and state-of-the-art methodology to provide solutions to analysts, policy makers & regulators.

REFERENCES

- [1] S. Dewan and H. Mendelson, "Information technology and time-based competition in financial markets," *Management Science*, vol. 44, no. 5, pp. 595–609, May 1998. <https://doi.org/10.1287/mnsc.44.5.595>
- [2] A. Cowles 3rd, "Can stock market forecasters forecast?" *Econometrica*, vol. 1, no. 3, pp. 309–324, Jul. 1933. <https://doi.org/10.2307/1907042>
- [3] O. V. Groos and A. Pritchard, "Documentation Notes," *Journal of Documentation*, vol. 25, no. 4, pp. 344–349, 1969. <https://doi.org/10.1108/eb026482>
- [4] F. Blanco-Mesa, J. M. Merigó, and A. M. Gil-Lafuente, "Fuzzy decision making: A bibliometric-based review," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 3, pp. 2033–2050, 2017. <https://doi.org/10.3233/JIFS-161640>
- [5] F. Black, "Noise," *The Journal of Finance*, vol. 41, no. 3, pp. 528–543, Jul. 1986. <https://doi.org/10.1111/j.1540-6261.1986.tb04513.x>
- [6] T. Lux and M. Marchesi, "Scaling and criticality in a stochastic multi-agent model of a financial market," *Nature*, vol. 397, no. 6719, pp. 498–500, 1999. <https://doi.org/10.1038/17290>
- [7] S. Chottiner, "Stock Market Research Methodology: A Case for the Systems Approach," *Decision Sciences*, vol. 3, no. 2, pp. 45–53. <https://doi.org/10.1111/j.1540-5915.1972.tb00535.x>
- [8] G. Coyle, "Qualitative and quantitative modelling in system dynamics: Some research questions," *System Dynamics Review*, vol. 16, no. 3, pp. 225–244, 2000. [https://doi.org/10.1002/1099-1727\(200023\)16:3<225::AID-SDR195>3.0.CO;2-D](https://doi.org/10.1002/1099-1727(200023)16:3<225::AID-SDR195>3.0.CO;2-D)
- [9] J. Hansen, "Technical market analysis using a computer," in *Proceedings of the 1956 11th ACM national meeting*, ACM, 1956, pp. 37–40. <https://doi.org/10.1145/800258.808943>
- [10] R. A. Levy, "Conceptual foundations of technical analysis," *Financial Analysts Journal*, vol. 22, no. 4, pp. 83–89, 1966. <https://doi.org/10.2469/faj.v22.n4.83>
- [11] J. Felsen, "Learning pattern recognition techniques applied to stock market forecasting," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 5, no. 6, pp. 583–594, Nov. 1975. <https://doi.org/10.1109/TSMC.1975.4309399>
- [12] E. I. Altman, "Statistical classification models applied to common stock analysis," *Journal of Business Research*, vol. 9, no. 2, pp. 123–149, Jun. 1981. [https://doi.org/10.1016/0148-2963\(81\)90001-1](https://doi.org/10.1016/0148-2963(81)90001-1)
- [13] M. C. Spooner, "Origin of fundamental analysis," *Financial Analysts Journal*, vol. 40, no. 4, pp. 79–80, 1984. <https://doi.org/10.2469/faj.v40.n4.79>
- [14] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," *The Journal of Finance*, vol. 55, no. 4, pp. 1705–1765, Aug. 2000. <https://doi.org/10.1111/0022-1082.00265>
- [15] M. Lam, "Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis," *Decision Support Systems*, vol. 37, no. 4, pp. 567–581, Sep. 2004. [https://doi.org/10.1016/S0167-9236\(03\)00088-5](https://doi.org/10.1016/S0167-9236(03)00088-5)
- [16] M. Paliwal, and U. A. Kumar, "Neural networks and statistical techniques: A review of applications," *Expert Systems with Applications*, vol. 36, no. 1, pp. 2–17, Jan. 2009. <https://doi.org/10.1016/j.eswa.2007.10.005>
- [17] C. Jiang, K. Liang, H. Chen, and Y. Ding, "Analyzing market performance via social media: A case study of a banking industry crisis," *Science China Information Sciences*, vol. 57, no. 5, pp. 1–18, 2014. <https://doi.org/10.1007/s11432-013-4860-3>
- [18] S. Mullainathan and J. Spiess, "Machine learning: An applied econometric approach," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 87–106, 2017. <https://doi.org/10.1257/jep.31.2.87>
- [19] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010. <https://doi.org/10.1007/s11192-009-0146-3>
- [20] K.-J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1–2, pp. 307–319, Sep. 2003. [https://doi.org/10.1016/S0925-2312\(03\)00372-2](https://doi.org/10.1016/S0925-2312(03)00372-2)
- [21] P. B. Henry, "Stock market liberalization, economic reform, and emerging market equity prices," *The Journal of Finance*, vol. 55, no. 2, pp. 529–564, Apr. 2000. <https://doi.org/10.1111/0022-1082.00219>
- [22] Y. Kara, M. A. Boyacioglu, and Ö. K. Baykan, "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*, vol. 38, no. 5, pp. 5311–5319, May 2011. <https://doi.org/10.1016/j.eswa.2010.10.027>
- [23] E. Guresen, G. Kayakutlu and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10389–10397, Aug. 2011. <https://doi.org/10.1016/j.eswa.2011.02.068>
- [24] M. T. Leung, H. Daouk, and A.-S. Chen, "Forecasting stock indices: A comparison of classification and level estimation models," *International Journal of Forecasting*, vol. 16, no. 2, pp. 173–190, Apr.–Jun. 2000. [https://doi.org/10.1016/S0169-2070\(99\)00048-5](https://doi.org/10.1016/S0169-2070(99)00048-5)
- [25] Y. Zhang and L. Wu, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8849–8854, Jul. 2009. <https://doi.org/10.1016/j.eswa.2008.11.028>
- [26] M. A. Boyacioglu and D. Avci, "An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7908–7912, Dec. 2010. <https://doi.org/10.1016/j.eswa.2010.04.045>
- [27] W. Leigh, R. Purvis, and J. M. Ragusa, "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support," *Decision Support Systems*, vol. 32, no. 4, pp. 361–377, Mar. 2002. [https://doi.org/10.1016/S0167-9236\(01\)00121-X](https://doi.org/10.1016/S0167-9236(01)00121-X)
- [28] P.-C. Chang and C.-H. Liu, "A TSK type fuzzy rule based system for stock price prediction," *Expert Systems with Applications*, vol. 34, no. 1, pp. 135–144, Jan. 2008. <https://doi.org/10.1016/j.eswa.2006.08.020>
- [29] G. Armano, M. Marchesi, and A. Murru, "A hybrid genetic-neural architecture for stock indexes forecasting," *Information Sciences*, vol. 170, no. 1, pp. 3–33, Feb. 2005. <https://doi.org/10.1016/j.ins.2003.03.023>
- [30] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques – Part II: Soft computing methods," *Expert Systems with Applications*, vol. 36, no. 3, part 2, pp. 5932–5941, Apr. 2009. <https://doi.org/10.1016/j.eswa.2008.07.006>
- [31] T. Ansari, M. Kumar, A. Shukla, J. Dhar, and R. Tiwari, "Sequential combination of statistics, econometrics and adaptive neural-fuzzy interface for stock market prediction," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5116–5125, Jul. 2010. <https://doi.org/10.1016/j.eswa.2009.12.083>
- [32] S. H. Kim and S. H. Chun, "Graded forecasting using an array of bipolar predictions: Application of probabilistic neural networks to a stock market index," *International Journal of Forecasting*, vol. 14, no. 3, pp. 323–337, Sep. 1998. [https://doi.org/10.1016/S0169-2070\(98\)00003-X](https://doi.org/10.1016/S0169-2070(98)00003-X)
- [33] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, Jan. 2015. <https://doi.org/10.1016/j.eswa.2014.07.040>
- [34] K. S. Kannan, P. S. Sekar, M. M. Sathik, and P. Arumugam, "Financial stock market forecast using data mining techniques," in *International Multiconference of Engineers and Computer Scientists*, 2010, pp. 555–559.
- [35] G. Dutta, P. Jha, A. K. Laha, and N. Mohan, "Artificial neural network models for forecasting stock price index in the Bombay stock exchange," *Journal of Emerging Market Finance*, vol. 5, no. 3, pp. 283–295, Dec. 2006. <https://doi.org/10.1177/097265270600500305>
- [36] I. Verma, L. Dey, and H. Meisheri, "Detecting, quantifying and accessing impact of news events on Indian stock indices," in *16th IEEE/WIC/ACM International Conference on Web Intelligence*, ACM, 2017, pp. 550–557. <https://doi.org/10.1145/3106426.3106482>
- [37] S. K. Khatri, H. Singhal, and P. Johri, "Sentiment analysis to predict Bombay stock exchange using artificial neural network," in *3rd International Conference on Reliability, Infocom Technologies and Optimization*, IEEE, 2014. <https://doi.org/10.1109/ICRITO.2014.7014714>
- [38] S. Deng, Z. J. Huang, A. P. Sinha, and H. Zhao, "The interaction between microblog sentiment and stock return: An empirical examination," *MIS Quarterly*, vol. 42, no. 3, pp. 895–918, 2018. <https://doi.org/10.25300/MISQ/2018/14268>
- [39] J. R. Piñero-Chousa, M. Á. López-Cabarcos, and A. M. Pérez-Pico, "Examining the influence of stock market variables on microblogging sentiment," *Journal of Business Research*, vol. 69, no. 6, pp. 2087–2092, Jun. 2016. <https://doi.org/10.1016/j.jbusres.2015.12.013>
- [40] L. Kristoufek, "Can Google Trends search queries contribute to risk diversification?" *Scientific Reports*, vol. 3, Article number 2713, 2013. <https://doi.org/10.1038/srep02713>

- [41] S. Agarwal, S. Kumar, and U. Goel, "Stock market response to information diffusion through internet sources: A literature review," *International Journal of Information Management*, vol. 45, pp. 118–131, Apr. 2019. <https://doi.org/10.1016/j.ijinfomgt.2018.11.002>
- [42] Y. Iyanar and R. Prasad, "Impact of CSR activities on shareholders' wealth in Indian companies," in *2018 International Conference on Advances in Computing, Communications and Informatics*, IEEE, 2018, pp. 2196–2199. <https://doi.org/10.1109/ICACCI.2018.8554711>
- [43] S. X. Xu and X. Zhang, "Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction," *MIS Quarterly*, vol. 37, no. 4, pp. 1043–1068, Dec. 2013. <https://doi.org/10.25300/MISQ/2013/37.4.03>
- [44] K. Hoang, D. Cannavan, R. Huang, and X. Peng, "Predicting stock returns with implied cost of capital: A partial least squares approach," *Journal of Financial Markets*, article number 100576, 2020, in press. <https://doi.org/10.1016/j.finmar.2020.100576>
- [45] T. Arshinova, "Construction of equity portfolio on the basis of data envelopment analysis approach," *Applied Computer Syst.*, vol. 45, no. 1, pp. 104–108, Dec. 2011. <https://doi.org/10.2478/v10143-011-0050-1>
- [46] R. K. Raut and R. Kumar, "Investment decision-making process between different groups of investors: A study of Indian stock market," *Asia-Pacific Journal of Management Research and Innovation*, vol. 14, no. 1–2, pp. 39–49, Mar. & Jun. 2018. <https://doi.org/10.1177/2319510X18813770>
- [47] V. P. Ramesh, P. Baskaran, A. Krishnamoorthy, D. Damodaran, and P. Sadasivam, "Back propagation neural network based big data analytics for a stock market challenge," *Communications in Statistics - Theory and Methods*, vol. 48, no. 14, pp. 3622–3642, 2019. <https://doi.org/10.1080/03610926.2018.1478103>
- [48] R. Dash and P. K. Dash, "A hybrid stock trading framework integrating technical analysis with machine learning techniques," *The Journal of Finance and Data Science*, vol. 2, no. 1, pp. 42–57, Mar. 2016. <https://doi.org/10.1016/j.jfds.2016.03.002>
- [49] M. R. Senapati, S. Das, and S. Mishra, "A novel model for stock price prediction using hybrid neural network," *Journal of The Institution of Engineers (India): Series B*, vol. 99, no. 6, pp. 555–563, Dec. 2018. <https://doi.org/10.1007/s40031-018-0343-7>
- [50] R. Arjun and K. R. Suprabha, "Forecasting banking sectors in Indian stock markets using machine intelligence," *International Journal of Hybrid Intelligent Systems*, vol. 15, no. 3, pp. 129–142, 2019. <https://doi.org/10.3233/HIS-190266>
- [51] L. Khansa and D. Liginlal, "Predicting stock market returns from malicious attacks: A comparative analysis of vector autoregression and time-delayed neural networks," *Decision Support Systems*, vol. 51, no. 4, pp. 745–759, Nov. 2011. <https://doi.org/10.1016/j.dss.2011.01.010>
- [52] R. Bisoi and P. K. Dash, "A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter," *Applied Soft Computing*, vol. 19, pp. 41–56, Jun. 2014. <https://doi.org/10.1016/j.asoc.2014.01.039>
- [53] F. Akhtar, K. S. Thyagaraj, and N. Das, "The impact of social influence on the relationship between personality traits and perceived investment performance of individual investors: Evidence from Indian stock market," *International Journal of Managerial Finance*, vol. 14, no. 1, pp. 130–148, 2018. <https://doi.org/10.1108/IJMF-05-2016-0102>
- [54] A. Abraham, B. Nath, and P. K. Mahanti, "Hybrid intelligent systems for stock market analysis," in Alexandrov V. N., Dongarra J. J., Juliano B. A., Renner R. S., Tan C. J. K. (eds) *Computational Science - ICCS 2001*. ICCS 2001. Lecture Notes in Computer Science, vol 2074. Springer, Berlin, Heidelberg, 2001. https://doi.org/10.1007/3-540-45718-6_38
- [55] A. Goyal and I. Welch, "Predicting the equity premium with dividend ratios," *Management Science*, vol. 49, no. 5, pp. 639–654, May 2003. <https://doi.org/10.1287/mnsc.49.5.639.15149>
- [56] T. Zorn, D. Dudley, and B. Jirasakuldech, "P/E changes: Some new results," *Journal of Forecasting*, vol. 28, no. 4, pp. 358–370, Jul. 2009. <https://doi.org/10.1002/for.1097>
- [57] J.-L. Wu and Y.-H. Hu, "Price-dividend ratios and stock price predictability," *Journal of Forecasting*, vol. 31, no. 5, pp. 423–442, Aug. 2012. <https://doi.org/10.1002/for.1231>
- [58] H. Allen and M. P. Taylor, "Charts, noise and fundamentals in the London foreign exchange market," *The Economic Journal*, vol. 100, no. 400, pp. 49–59, Apr. 1990. <https://doi.org/10.2307/2234183>
- [59] G. Baltussen, S. van Bakkum, and Z. Da, "Indexing and stock market serial dependence around the world," *Journal of Financial Economics*, vol. 132, no. 1, pp. 26–48, Apr. 2019. <https://doi.org/10.1016/j.jfineco.2018.07.016>
- [60] M. A. Ferreira and P. Santa-Clara, "Forecasting stock market returns: The sum of the parts is more than the whole," *Journal of Financial Economics*, vol. 100, no. 3, pp. 514–537, Jun. 2011. <https://doi.org/10.1016/j.jfineco.2011.02.003>
- [61] Y. Gorodnichenko and M. Weber, "Are sticky prices costly? Evidence from the stock market," *American Economic Review*, vol. 106, no. 1, pp. 165–199, Jan. 2016. <https://doi.org/10.1257/aer.2013.1513>
- [62] J. Greenwood and B. Jovanovic, "The information-technology revolution and the stock market," *American Economic Review*, vol. 89, no. 2, pp. 116–122, May 1999. <https://doi.org/10.1257/aer.89.2.116>
- [63] B. Hobijn and B. Jovanovic, "The information-technology revolution and the stock market: Evidence," *The American Economic Review*, vol. 91, no. 5, pp. 1203–1220, Dec. 2001. <https://doi.org/10.1257/aer.91.5.1203>
- [64] J. Laitner and D. Stolyarov, "Technological change and the stock market," *American Economic Review*, vol. 93, no. 4, pp. 1240–1267, Sep. 2003. <https://doi.org/10.1257/000282803769206287>
- [65] D. C. Parkes and M. P. Wellman, "Economic reasoning and artificial intelligence," *Science*, vol. 349, no. 6245, pp. 267–272, Jul. 2015. <https://doi.org/10.1126/science.aaa8403>
- [66] S. Sudhakaran and P. Balasubramanian, "A study on the impact of macroeconomic factors on S&P BSE Bankex returns," in *2016 International Conference on Advances in Computing, Communications and Informatics*, IEEE, 2016, pp. 2614–2618. <https://doi.org/10.1109/ICACCI.2016.7732452>
- [67] B. Nikita, P. Balasubramanian, and L. Yermal, "Impact of key macroeconomic variables of India and USA on movement of the Indian stock return in case of S&P CNX Nifty," in *2017 International Conference on Data Management, Analytics and Innovation*, IEEE, 2017, pp. 330–333. <https://doi.org/10.1109/ICDMAI.2017.8073536>
- [68] P. Krishnamurthy, P. Balasubramanian, and D. Mohan, "Study on relationship between exchange rate return and various stock indices returns," in *2017 International Conference on Data Management, Analytics and Innovation*, IEEE, 2017, pp. 316–320. <https://doi.org/10.1109/ICDMAI.2017.8073533>
- [69] I. Zheludev, R. Smith, and T. Aste, "When can social media lead financial markets?" *Scientific Reports*, vol. 4, Article number 4213, 2014. <https://doi.org/10.1038/srep04213>
- [70] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, Mar. 2011. <https://doi.org/10.1016/j.jocs.2010.12.007>
- [71] T. Preis, H. S. Moat, and H. E. Stanley, "Quantifying trading behavior in financial markets using Google Trends," *Scientific Reports*, vol. 3, Article number 1684, 2013. <https://doi.org/10.1038/srep01684>
- [72] F. Nagle, "Stock market prediction via social media: The importance of competitors," *Academy of Management Proc.*, 2013. Retrieved from <https://journals.aom.org/doi/abs/10.5465/ambpp.2013.17557abstract>
- [73] M. Nardo, M. Petracco-Giudici, and M. Naltsidis, "Walking down Wall Street with a tablet: A survey of stock market predictions using the web," *Journal of Economic Surveys*, vol. 30, no. 2, pp. 356–369, Apr. 2016. <https://doi.org/10.1111/joes.12102>
- [74] P. Saxena, B. Pant, R. H. Goudar, S. Srivastav, V. Garg, and S. Pareek, "Future predictions in Indian stock market through linguistic-temporal approach," in *7th International Conference on Intelligent Systems and Control*, IEEE, 2013, pp. 416–420. <https://doi.org/10.1109/ISCO.2013.6481191>
- [75] M. Alanyali, H. S. Moat, and T. Preis, "Quantifying the relationship between financial news and the stock market," *Scientific Reports*, vol. 3, article number 3578, 2013. <https://doi.org/10.1038/srep03578>
- [76] T. Geva and J. Zahavi, "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news," *Decision Support Systems*, vol. 57, pp. 212–223, Jan. 2014. <https://doi.org/10.1016/j.dss.2013.09.013>
- [77] K. Nam and N. Seong, "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market," *Decision Support Systems*, vol. 117, pp. 100–112, Feb. 2019. <https://doi.org/10.1016/j.dss.2018.11.004>
- [78] R. Dasgupta and R. Singh, "Investor sentiment antecedents: A structural equation modeling approach in an emerging market context," *Review of Behavioral Finance*, vol. 11, no. 1, pp. 36–54, 2018. <https://doi.org/10.1108/RBF-07-2017-0068>
- [79] D. Kinslin and V. P. Velmurugan, "Investors' behavior and perceptions towards stock market: Structural equation modeling approach," *International Journal of Engineering & Technology*, vol. 7, no. 4.36, pp. 586–591, 2018. <https://doi.org/10.14419/ijet.v7i4.36.24205>

- [80] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "Predicting stock market price movement using sentiment analysis: Evidence from Ghana," *Applied Computer Systems*, vol. 25, no. 1, pp. 33–42, May 2020. <https://doi.org/10.2478/acss-2020-0004>
- [81] A. Al-Nasseri and F. Menla Ali, "What does investors' online divergence of opinion tell us about stock returns and trading volume?" *Journal of Business Research*, vol. 86, pp. 166–178, May 2018. <https://doi.org/10.1016/j.jbusres.2018.01.006>
- [82] C. Antoniou, J. A. Doukas, and A. Subrahmanyam, "Investor sentiment, beta, and the cost of equity capital," *Management Science*, vol. 62, no. 2, pp. 347–367, Feb. 2016. <https://doi.org/10.1287/mnsc.2014.2101>
- [83] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Reviews of Modern Physics*, vol. 81, no. 2, pp. 591–646, Apr.–Jun. 2009. <https://doi.org/10.1103/RevModPhys.81.591>
- [84] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann, "Noise trader risk in financial markets," *Journal of Political Economy*, vol. 98, no. 4, pp. 703–738, Aug. 1990. <https://doi.org/10.1086/261703>
- [85] O. Altunkılıç, V. S. Balashov, and R. S. Hansen, "Are analysts' forecasts informative to the general public?" *Management Science*, vol. 59, no. 11, pp. 2550–2565, Nov. 2013. <https://doi.org/10.1287/mnsc.2013.1721>
- [86] B. G. Deshmukh, P. S. Jain, M. S. Patwardhan, and V. Kulkarni, "Spin-offs in Indian stock market owing to Twitter sentiments, commodity prices and analyst recommendations," in *2016 International Conference on Advances in Information Communication Technology and Computing*, ACM, Article No. 77, 2016. <https://doi.org/10.1145/2979779.2979856>
- [87] P. H. Cootner (Ed.), *The Random Character of Stock Market Prices*. The MIT Press, 1967.

Arjun Remadevi Somanathan is a PhD candidate at the NITK, Surathkal. He holds an Masters in Software Engineering from CUSAT, India. He has published research and also reviews for refereed journal/conferences. Main interests are information systems, management science and applied research. Email: arjrs123@gmail.com
ORCID: <http://orcid.org/0000-0002-2770-6164>

Dr Suprabha Kudigrama Rama is an Assistant Professor at the NITK, India. She has obtained PhD from VTU, India. She has guided research scholars and published in journals and conferences while secured best paper awards. She is an investigator for funded research projects. Email: suprabha.kr@gmail.com

APPENDIX I

STUDIES RELATED TO FINANCIAL AND ECONOMIC FACTORS

Financial factors	[54]; [55]; [56]; [57]
Economic factors (micro/macro level)	[58]; [59]; [60], [61]; [62]; [63]; [64]; [65]; [66]; [67]; [68]; [69]

APPENDIX II

STUDIES RELATED TO SOCIAL AND ORGANIZATIONAL FACTORS

Twitter	[69]; [70]
Google trends	[40]; [71]
Other social media factors	[17]; [72]; [73]; [74]; [75]
Financial news diffusion	[36]; [76]; [77];
Investor sentiment	[37]; [38]; [39]; [78]; [80]; [81]; [82]
Peer influence	[53]; [79]
Group dynamics	[46]; [81]; [83]; [84]
Analysts' view	[85]; [86]; [87]