Grey Exponential Smoothing and Sentiment Classification Approach for Stock Prediction Using Big Data Analytics: A Case Study on Colombo Stock Exchange, Sri Lanka

D. M. K. N. Seneviratna[1], M.V.D.H.P Malawana[2], R.M.K.T. Rathnayaka[3]

1 Department of Interdisciplinary Studies, Faculty of Engineering, University of Ruhuna, Galle, Sri Lanka.
2 Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka
3Department of Physical Sciences and Technology, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

seneviratna@is.ruh.ac.lk, kapiar@appsc.sab.ac.lk

Abstract*—*

Sentiment Analysis is an innovative development in machine learning with the use of natural language processing techniques to derive emotions as positive, negative, neutral from public opinions in information. The main objective of this study is to introduce a sentiment analysis mechanism using social media data within big data distributed environment for stock market predictions. The proposed methodology is validated for the Colombo Stock market during the period from 2010 January to 2020 based on the Twitter feed content and stock market movements. Based on the past tweets, the current research examines the effectiveness of various machine learning techniques such as K Nearest Neighbour, Decision Tree Model, Support Vector Machine, Grey Exponential Smoothing model, and Multinomial Naïve Bayes machine learning to predict stock market indices. The mentioned algorithms were trained and tested through 80% of the data was used for training and 20% was used for testing. The key finding of this research suggested that the Grey Exponential Smoothing model and Naïve Bayes perform well in sentiment classification. Furthermore, results confirmed that the public sentiment highly influences the market fluctuations and economic elements like monetizing policies, government changes, unexpected pandemics, interest rates, the confidence of expectations of the public, confidence in economic growth are significantly make sense on the stock market.

Keywords*—* **Stock Market, Sentiment Analysis, Machine Learning**

REFERENCES

- R. M. K. T. Rathnayaka, D. M. K. . Seneviratna, and W. Jianguo, "Grey system based novel approach for stock market forecasting," Grey Syst. Theory Appl., vol. 5, no. 2, pp. 178–193, 2015, doi: 10.1108/gs-04-2015-0014.

- Asur, S. and Huberman, B. A. (2010) 'Predicting the future with social media', Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010, 1, pp. 492–499. doi: 10.1109/WI-IAT.2010.63.

- D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," Int. J. Financ. Stud., vol. 7, no. 2, 2019, doi: 10.3390/ijfs7020026.

- M. Skuza and A. Romanowski, "Sentiment analysis of Twitter data within big data distributed environment for stock prediction," Proc. 2015 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2015, vol. 5, pp. 1349–1354, 2015, doi: 10.15439/2015F230.

- S. K. Khatri and A. Srivastava, "Using sentimental analysis in prediction of stock market investment," 2016 5th Int. Conf. Reliab. Infocom Technol. Optim. ICRITO 2016 Trends Futur. Dir., pp. 566–569, 2016, doi: 10.1109/ICRITO.2016.7785019.

- Parveen, H. and Pandey, S. (2017) 'Sentiment analysis on Twitter Data-set using Naive Bayes algorithm', Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016. IEEE, pp. 416–419. doi: 10.1109/ICATCCT.2016.7912034.

- Zainuddin, N. and Selamat, A. (2014) 'Sentiment analysis using Support Vector Machine', I4CT 2014 - 1st International Conference on Computer, Communications, and Control Technology, Proceedings, (May 2016), pp. 333–337. doi: 10.1109/I4CT.2014.6914200.

- Bollen, J., Mao, H. and Zeng, X. (2011) 'Twitter mood predicts the stock market', Journal of Computational Science, 2(1), pp. 1–8. doi: 10.1016/j.jocs.2010.12.007.

- Dickinson, B. and Hu, W. (2015) 'Sentiment Analysis of Investor Opinions on Twitter', Social Networking, 04(03), pp. 62–71. doi: 10.4236/sn.2015.43008.

- Elangkumaran, P. and Jenitta, J. . (2014) 'Macroeconomic Variables and Stock Prices : A Study of Colombo Stock Exchange (CSE) in Sri Lanka', International Journal of Research in Commerce, IT & Management, 4(January). doi: 2231-5756.

- Bhardwaj, Y. Narayan, Vanraj, Pawan, and M. Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty," Procedia Comput. Sci., vol. 70, pp. 85–91, 2015, doi: 10.1016/j.procs.2015.10.043.

- Zhao, Y. He, C. Yuan, and Y. Huang, "Stock market prediction exploiting microblog sentiment analysis," Proc. Int. Jt. Conf. Neural Networks, vol. 2016-Octob, no. 1, pp. 4482–4488, 2016, doi: 10.1109/IJCNN.2016.7727786.

- S. Pacharkar, P. Kulkarni, Y. Mishra, A. Jagadambe, and S. G. Shaikh, "Predicting Stock Market Investment Using Sentiment Analysis," Int. J. Adv. Res. Comput. Commun. Eng. ISO, vol. 3297, pp. 109–114, 2007, doi: 10.17148/IJARCCE.2018.7321.

- J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," J. Comput. Sci., vol. 2, no. 1, pp. 1–8, 2011, doi: 10.1016/j.jocs.2010.12.007.

- U. Sekaran and R. Bougie, "Research Methods for Bussiness A Skill-Bulding Approach," pp. 1–447, 2016.