

Table 3. Performance metrics for various algorithms

Algorithm	MAE	MAPE	RMSE	Accuracy (%)
Random forest	1.126	1.578	1.463	67.32
Gradient boost	1.098	1.549	1.226	68.67
XG-Boost	0.985	1.395	1.124	71.78

6. CONCLUSION & FUTURE WORK

In this research, the sentiment of comments and tweets were also considered to improve the stock prediction accuracy by introducing the numerical value derived from sentiment analysis along with the data used for predicting the stock trend. In this work, the experimentation is carried out on the data collected from YouTube. The work can be further extended by performing the experiments on data collected from different sources and deploying the stacking concept for obtaining the accurate prediction. The training data need to be improved by excluding irrelevant comments and the amount of videos to analyze need to be increased. Depending on the model and the data selected, the boosting or bagging approaches need to be applied which is used to overcome over-fitting problem but boosting generally exhibits better bias in case of low performance individual models. Depending on the model and the data selected, the boosting or bagging approaches need to be applied which is used to overcome over-fitting problem but boosting generally exhibits better bias in case of low performance individual models. GBM performs better than RF if parameters tuned carefully, Gradient Boosting builds trees in the order of one at a time, where each new tree helps to rectify errors made in the previously trained tree. There exists a correlation among the likes count, dislikes count, comments on YouTube and tweets from twitter but as the variation is very high it is not possible to predict accurately using individual like or dislike counts. In this research, the sentiment of comments and tweets were also considered to improve the stock prediction accuracy by answering the research question, “*Can the day to day difference of likes count and dislikes count and comments on a YouTube video and tweets from the twitter be used to determine the respective stock trend prediction*” by introducing the numerical value derived from sentiment analysis is used along with the data used for predicting the stock trend. In this work, the experimentation is carried out on the data collected from YouTube. The work can be further extended by performing the experiments on data collected from different sources and deploying the stacking concept for obtaining the accurate prediction. The training data need to be improved by excluding irrelevant comments and the amount of videos to analyze need to be increased.

REFERENCES

- [1] Bhardwaj, A., Narayan, Y., Vanraj, Pawan, Dutta, M. (2015). Sentiment analysis for Indian stock market prediction using Sensex and nifty. *Procedia Computer Science*, 70: 85-91. <https://doi.org/10.1016/j.procs.2015.10.043>
- [2] Bilinski, P. (2019). Living in the digital age: The usefulness of Twitter, YouTube and Instagram in corporate communication. *Youtube and Instagram in Corporate Communication*. <http://dx.doi.org/10.2139/ssrn.3311183>
- [3] Kramer, M. (2019). Social media networking and the impact it has on enterprise application software. *Global Journal of Business Research*, 13(2): 37-45.
- [4] Bhardwaj, N., Ansari, M.A. (2019). Prediction of stock market using machine learning algorithms. *International Research Journal of Engineering and Technology*, 6(5).
- [5] Basak, S., Kar, S., Saha, S., Khaidem, L., Dey, S.R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47: 552-567. <https://doi.org/10.1016/j.najef.2018.06.013>
- [6] Martin, H., Teodor, I. (2018). Predicting like-ratio on YouTube videos using sentiment analysis on comments. Thesis, School of Electrical Engineering and Computer Science (EECS).
- [7] Rahim, M.S., Chowdhury, A.Z.M.E., Islam, M.A., Islam, M.R. (2017). Mining trailers data from Youtube for predicting gross income of movies. In 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 551-554. <https://doi.org/10.1109/R10-HTC.2017.8289020>
- [8] Golmohammadi, K., Zaiane, O.R. (2017). Sentiment analysis on Twitter to improve time series contextual anomaly detection for detecting stock market manipulation. In *International Conference on Big Data Analytics and Knowledge Discovery*, pp. 327-342. https://doi.org/10.1007/978-3-319-64283-3_24
- [9] Sharief, P.N., Ali, M.M., Prasad, V. (2017). Forecasting stock prices of selected FMCG companies listed in NSE India limited using technical analysis. *International Journal of Applied Financial Management Perspectives*.
- [10] Bhuiyan, H., Ara, J., Bardhan, R., Islam, M.R. (2017). Retrieving YouTube video by sentiment analysis on user comment. *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 474-478. <https://doi.org/10.1109/ICSIPA.2017.8120658>
- [11] Amar, K. (2013). Polarity trend analysis of public sentiment on YouTube. In *Proceedings of the 19th International Conference on Management of Data*, pp. 125-128. <https://dl.acm.org/doi/abs/10.5555/2694476.2694505>
- [12] Deng, S., Mitsubuchi, T., Shioda, K., Shimada, T., Sakurai, A. (2011). Combining technical analysis with sentiment analysis for stock price prediction. In *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pp. 800-807. <https://doi.org/10.1109/DASC.2011.138>