output data with high accuracy and aggregated features, which is best source for further data processing applications. Here authors worked with data collected by limited source and fields only, but the same approach can be applied with multiple sources and fields. Further we will extend our work on more different types of data sources with parallel processing and system based on Hadoop and analyze the result.

References

[1] Bello-Orgaz, G.; Jung, J; Camacho, D. (2015). Social big data: Recent achievements and new challenges. Information Fusion, pp. 1-15.
[2] Taleb, I.; Dssouli, R.; Serhani, M.A. (2015): Big Data Pre-Processing: A Quality Framework. 2015 IEEE International Congress on Big Data, pp. 191-198.
[3] Arputhamary, B.; Arockiam, L. (2014): A Review on Big Data Integration. IJCA Proceedings on International Conference on Advanced Computing and Communication Techniques for High Performance Applications ICACCTHPA, pp. 21-26.
[4] Dong, X.L.; Srivastava, D. (2013): Big Data Integration. Proceedings of the 2013 IEEE International Conference on Data Engineering, pp. 1245-1248.
[5] Lomborg, S.; Bechmann, A. (2014): Using APIs for Data Collection on Social Media. The Information Society 30, pp. 256-265.
[6] Saranya, C.; Manikandan, G. (2013): A study on Normalization Techniques for Privacy Preserving Data Mining. International Journal of Engineering and Technology 5, pp. 2701-2704.
[7] García, S.; Ramírez-Gallego, S.; Luengo, J.; Benítez, J.M.; Herrera, F. (2016): Big data preprocessing: methods and prospects. Big Data Analytics 1(9), pp. 1-22.
[8] Jason Brownlee (2017) Python Machine Learning para. 5. https://machinelearningmastery.com/handle-missing-data-python/. Accessed: 10 Oct 2020.
[9] Huh, J.; Grundy, J.; Hosking, J.; Liu, K.; Amor, R. (2009): Integrated Data Mapping for a Software Meta-tool. Proceedings of the 2009 Australian Software Engineering Conference, ASWEC, pp. 111-120.
[10] Dai, H.; Zhang, S.; Wang, L.; Ding, Y. (2016): Research and Implementation of Big Data Preprocessing System Based on Hadoop. 2016 IEEE International Conference on Big Data Analysis (ICBDA), pp. 1-5.
[11] Zhou, J.; Hu, L.; Wang, F.; Lu, H.; Zhao, K. (2013): An Efficient Multidimensional Fusion Algorithm for IoT Data Based on Partitioning. Tsinghua Science and Technology 18(4), pp. 369-378.
[12] Arputhamary, B.; Arockiam, L. (2015): Data Integration in Big Data Environment. Bonfring International Journal of Data Mining 5: pp. 1 - 5.
[13] Shalabi, L.A.; Shaaban, Z. (2006): Normalization as a Preprocessing Engine for Data Mining and the Approach of Preference Matrix. International Conference on Dependability of Computer Systems, pp. 207-214.
[14] Bhadani, A.K.; Jothimani, D. (2016): Big Data: Challenges, Opportunities and Realities. Effective Big Data Management and Opportunities for Implementation, Pennsylvania, USA, IGI Global, pp. 1-24.
[15] Martinchek (2016). 2012-2016 Facebook Posts Retrieved from https://data.world/martinchek/2012-2016-facebook-posts
[16] Chaudhary, A.; Kolhe, S.; Kamal, R. (2016): A hybrid ensemble for classification in multiclass datasets: An application to oilseed disease dataset. Computers and Electronics in Agriculture 124, pp. 65–72.
[17] Chaudhary, A.; Kolhe, S.; Kamal, R. (2016): An improved random forest classifier for multi-class classification. Information Processing in Agriculture 3, pp. 215–222.
[18] Chaudhary, A.; Kolhe, S.; Kamal, R. (2013): Machine learning classification techniques: A comparative study. International Journal on Advanced Computer Theory and Engineering 2(4), pp. 21–25.
[19] Chaudhary, A.; Kolhe, S.; Kamal, R. (2013): Machine Learning Techniques for Mobile Devices: A Review. International Journal of Engineering Research and Applications 3(6), pp. 913–917.
[20] Jadiya, A.K.; Thakur, R. (2019): Efficient Workflow for Social Big Data Processing. Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA).