Modeling of class imbalance handling with optimal deep learning enabled big data classification model

Article type: Research Article

Authors: S, Varshavardhini (https://content.iospress.com:443/search?q=author%3A%28%22S, Varshavardhini%22%29)* | A, Rajesh (https://content.iospress.com:443/search?q=author%3A%28%22A, Rajesh%22%29)

Affiliations: Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India

Correspondence: [*] Corresponding author: Varshavardhini S, Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India. E-mail: varshavardhinis63@gmail.com (mailto:varshavardhinis63@gmail.com).

Abstract: Big data is the amount of data that surpasses the ability to process the data of a system concerning memory usage and computation time. It is commonly applied in several domains like healthcare, education, social networks, e-commerce, etc., as they have progressively obtained a massive quantity of input data. A major research problem is big data analytics, which can be carried out using expert systems and deep structured architectures. Besides, data wrangling and class imbalance data handling are challenging issues that need to be resolved in big data analytics. Class imbalance data degrade the performance of the classification model, which remains a challenging process due to the heterogeneous and complex structure of the comparatively huge datasets. Thus, the research focused on presenting a Class Imbalance Handling with Optimal Deep Learning Enabled Big Data Classification (CIHODL-BDC) framework. The core perception of the CIHODL-BDC framework helps to classify the big data in the Hadoop MapReduce framework. To accomplish this, the presented CIHODL-BDC model initially performs a data wrangling process is performed to alter the unrefined data into a useful layout. Next, the CIHODL-BDC model handles the class imbalance problem using a grey wolf optimizer (GWO) with Synthetic Minority Oversampling (SMOTE) technique. Besides, the Adam optimizer procedure with the Bidirectional Long Short Term Memory (BiLSTM) approach is performed to categorize the big data. The result analysis of the proposed CIHODL-BDC model is evaluated by two standard datasets. The simulation outcomes revealed the elevated performance of the CIHODL-BDC approach over existing methods.

Keywords: Data wrangling, big data analytics, hadoop mapreduce, class imbalance data handling, deep learning

DOI: 10.3233/IDT-230198

Journal: Intelligent Decision Technologies (https://content.iospress.com:443/journals/intelligent-decision-technologies), vol. 17, no. 4, pp. 1179-1197, 2023

Published: 20 November 2023

Price: EUR 27.50