

## On the Performance of Nature-Inspired Instance Selection Techniques Hybridized with Machine Learning Algorithms for Speed Optimization in Big Dataset Analysis

<sup>1</sup>Andronicus Ayobami A, <sup>2</sup>Arasomwan Martins

<sup>1</sup>Department of Computer Science and Informatics, University of the Free State, 9301 Bloemfontein South Africa

<sup>2</sup>School of Natural and Applied Sciences, Sol Plaatje University, Private Bag X5008, Kimberly, 8300

The volume of data generated daily through the use of Google search engine, Twitter, Instagram, Facebook, etc. has become overwhelming. Unfortunately, traditional data analytics techniques are fast losing their capabilities and efficiencies handling such volume of data. This challenge has motivated several researchers to design different efficient and robust methods to handle the analysis of big dataset. These techniques include data condensation, divide and conquer, density-based approaches, and distributed computing. Moreover, some of these techniques aim at reducing the volume of input dataset to speed up the computation time of big dataset analysis. Interestingly, faster and more accurate big dataset processing techniques can be developed using Machine Learning (ML) algorithms and Nature-inspired instance selection Techniques. This paper presents two hybrid Nature-Inspired ML-based methods for improving the computation speed of big data analytics. In the first method, we combine a cuckoo search based instance selection technique with four ML algorithms: Naïve Bayes, Random Forest, BayesNet and Artificial Neural Network. Besides, in the second method, we combine a flower pollination based instance selection technique with the four ML algorithms mentioned above. Moreover, we applied the combined methods to five large or medium-scale datasets, and the results show that they significantly improve the computational speed of big data analytics. Furthermore, the results show that Nature-Inspired instance selection techniques have strong data reduction capacity, and they can improve the training speed of ML algorithms, without significantly affecting their classification accuracy.

**Key Words:** Big Data Analytics, Machine Learning, Nature Inspired Algorithm, Instance Selection, data reduction.